(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 December 2002 (19.12.2002)

PCT

(10) International Publication Number
WO 2002/101357 A2

(51) International Patent Classification[7]: G01N

(21) International Application Number: PCT/US2002/018628

(22) International Filing Date: 10 June 2002 (10.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/297,277 10 June 2001 (10.06.2001) US

(71) Applicant *(for all designated States except US)*: IRM LLC [US/—]; c/o Sofia House, 48 Church Street, HM 12 Hamilton (BM).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: SU, Andrew, I. [US/US]; 8236 Caminito Modena, La Jolla, CA 92037 (US). HAMPTON, Garret, M. [IE/US]; 5659 Arbor Grove Court, San Diego, CA 92121 (US).

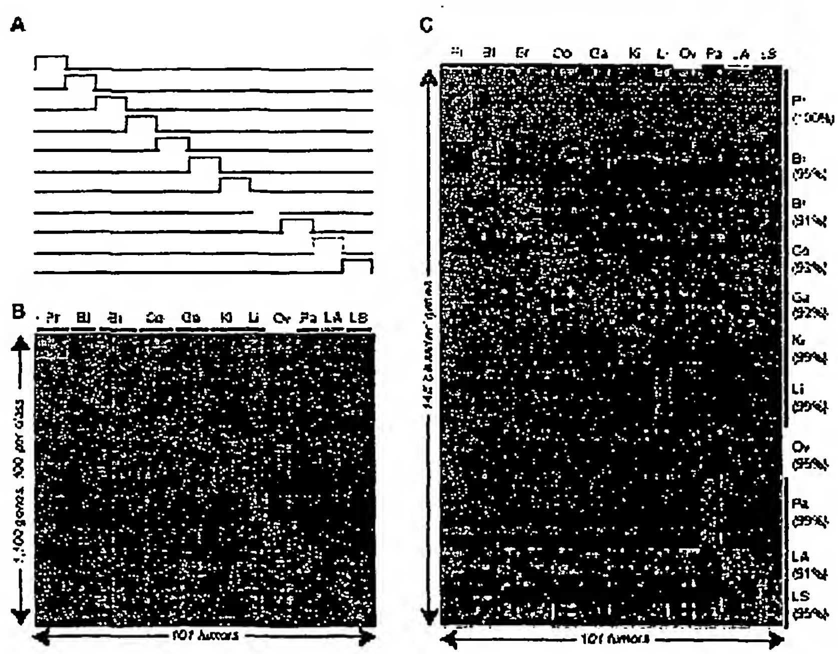
(74) Agent: SMITH, Timothy, L.; Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, Suite E225, San Diego, CA 92121 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: MOLECULAR SIGNATURES OF COMMONLY FATAL CARCINOMAS



(57) Abstract: This invention provides methods, kits, and algorithms for obtaining molecular signatures of cells based on their gene expression profiles. Devices for carrying out molecular signature analysis of unknown samples are also provided.

Published:
— *without international search report and to be republished upon receipt of that report*

(48) Date of publication of this corrected version: 12 February 2004

(15) Information about Correction: see PCT Gazette No. 07/2004 of 12 February 2004, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Molecular Signatures of Commonly Fatal Carcinomas

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. provisional application No. 60/297,277 filed June 10, 2001. The aforementioned application is incorporated herein by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] This invention pertains to the field of diagnosis, prognosis and treatment of carcinomas. In particular, the invention provides methods for identifying the anatomic origin of carcinomas.

Background

[0002] Cancer is a leading cause of death in the United States, causing one in four deaths, which is second only to heart disease. More than half a million people die of cancer each year in the United States. Four cancer sites, the lung, prostate, breast and colon, account for 56% of all new cancer cases and are the leading causes of cancer deaths for every racial and ethnic group, according to the *Annual Report to the Nation on the Status of Cancer*, 1973-1998 (see Howe et al., *J. Nat'l. Cancer Institute*, 93:824-842 (2001)).

[0003] In about 4% of all patients diagnosed with cancer, the observed tumor is due to metastasis and the primary tumor origin is undetermined (see Hillen, *Postgrad. Med. J.*, 76:690-693 (2000)). Thus, a central goal of cancer biology is the identification of molecules or sets of molecules that are unique to specific human carcinomas, both for the development of diagnostics and drugs for the treatment of disease, as well as ultimately to understand the mechanistic basis of tissue-specific tumorigenesis. Thus, the identification of genes whose expression is uniquely characteristic of tumors of diverse anatomic origins remains a central challenge to the development of new cancer therapies (see St Croix et al., *Science*, 289:1197-1202 (2000); Bittner et al., *Nature*, 406:536-540 (2000); Perou et al.,

Nature, 406:747-752 (2000); and Golub et al., *Science*, 286:531-537 (1999)). The present invention fulfills this and other needs.

SUMMARY OF THE INVENTION

[0004] The invention provides kits and methods for determining the origin of a tumor. In a first embodiment, the invention provides kits for identifying an origin of a tumor in a subject. These kits include: a) a probe that can detect an expression product of a gene in a first tumor class as indicated in Table 3; and b) a probe that can detect an expression product of a gene in a second tumor class as indicated in Table 3. The kits can also include additional probes, such two or more probes for each of the tumor classes, probes that are diagnostic for more than two tumor classes, or any combination thereof. In some embodiments, the kits include probes for at least one gene in each of at least ten tumor classes. The tumor classes for which the invention provides diagnostic kits include prostate cancer, breast cancer, colorectal cancer, lung adenocarcinoma, lung squamous cell carcinoma, ovarian cancer, gastroesophageal cancer, pancreatic cancer, liver cancer, kidney cancer and bladder cancer. The expression product that is detected can be, for example, an mRNA that is transcribed from the gene, a protein encoded by the gene, or a product of an enzymatic reaction catalyzed by a protein encoded by the gene.

[0005] Also provided by the invention are methods for identifying an origin of a tumor. These methods involve detecting in a tumor sample an expression level of at least two genes, each of which genes is diagnostic for a different tumor class as identified in Table 3. An elevated level expression for a gene indicates that the tumor originated from the tumor class for which the gene is diagnostic. The methods provided can be used to determine, for example, whether a tumor sample originated from a prostate cancer, breast cancer, colorectal cancer, lung adenocarcinoma, lung squamous cell carcinoma, ovarian cancer, gastroesophageal cancer, pancreatic cancer, liver cancer, kidney cancer or bladder cancer. In some embodiments, an expression level is determined for at least three genes, each of which genes is diagnostic for a different tumor class as identified in Table 3, for at least two genes that are both diagnostic for a single tumor class as identified in Table 3, or combinations thereof. For example, the invention provides methods in which an expression level is determined for at least two genes that are both diagnostic for a first tumor class as identified in Table 3, and at least two genes that are both diagnostic for a second tumor class as

identified in Table 3. The expression level of a gene in the tumor sample can be compared to the expression level of the gene in a non-cancer control sample, or to the expression level of the gene in a control sample obtained from a tumor of a different tumor class.

[0006] The invention also provides methods for identifying an origin of a tumor by: a) providing a predictor set that comprises expression levels for two or more genes, each of which is diagnostic for a different tumor class as identified in Table 3; b) detecting in a tumor sample an expression level of at one gene that is diagnostic for a tumor class as identified in Table 3; and c) calculating a vector distance from the expression level obtained from the tumor sample to each of the expression levels of the predictor set. The shortest vector distance from the unknown sample to one of the members of the predictor set indicates the origin of the tumor. In some embodiments, the predictor set includes expression levels for at least three genes, each of which genes is diagnostic for a different tumor class as identified in Table 3. The predictor set can include expression levels for at least two genes that are both diagnostic for a single tumor class as identified in Table 3. In some embodiments, the predictor set includes expression levels for at least two genes that are both diagnostic for a first tumor class as identified in Table 3, and at least two genes that are both diagnostic for a second tumor class as identified in Table 3. The predictor set, in some embodiments, includes expression levels for one or more genes in each of at least ten tumor classes identified in Table 3.

[0007] Methods for obtaining a predictor set for classifying a sample into one of two or more classes are also provided by the invention. These methods involve: a) obtaining a value for one or more features for each of a plurality of members of each of the classes; b) determining a Wilcoxon rank score for each of the features to eliminate nonpredictive features; and c) ranking the remaining features by predictive accuracy using a support vector machine. In some embodiments, the features are genes and the values are expression levels of the genes. The classes into which the sample is to be classified can include, for example, tumor classes, disease states, exposure to different conditions, and the like. The invention also provides computer-readable media and computers that are programmed to carry out the methods for obtaining a predictor set. These methods can further involve classifying a sample into one of the classes by: a) determining a value for one or more features in the sample; and b) calculating a vector distance from the obtained for the feature in the sample

to each of the expression levels of the predictor set, wherein the shortest vector distance indicates the class of which the sample is a member.

[0008] Methods for screening a subject for prostate cancer or at risk of developing prostate cancer are also provided by the invention. These methods involve:

a) detecting a level of expression of at least one gene in a sample of prostate tissue obtained from the subject to provide a first value, wherein the gene is selected from the group consisting of LIM, multidrug resistance-associated protein homolog (MRP4), T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I; and

b) comparing the first value with a level of expression of the gene in a sample of prostate tissue obtained from a disease-free subject, wherein a greater expression level in the subject sample compared to the sample from the disease-free subject is indicative of the subject having prostate cancer or at risk of developing prostate cancer.

[0009] The invention also provides methods for screening a subject for ovarian cancer or at risk of developing ovarian cancer. These methods involve:

a) detecting a level of expression of at least one gene in a sample of ovarian tissue obtained from the subject to provide a first value, wherein the gene is selected from the group consisting of laminin, alpha 5; vacuolar proton pump, beta polypeptide; putative cytoskeletal protein, natriuretic peptide receptor A, eyes absent homolog, U90916, AL049313, S100 alpha, keratinocyte transglutaminase, GPCR64, meis1, spondin 1, GPCR39, AL050069, mammoglobin 2, and branched chain aminotransferase 1, cytosolic; mesothelin and kallikrein 6; and

b) comparing the first value with a level of expression of the gene in a sample of ovarian tissue obtained from a disease-free subject, wherein a greater expression level in the subject sample compared to the sample from the disease-free subject is indicative of the subject having ovarian cancer or at risk of developing ovarian cancer.

[0010] The invention also provides methods for monitoring the progression of prostate cancer in a subject having, or at risk of having a prostate cancer. These methods involve measuring a level of expression of at least one gene selected from the group

consisting of LIM, multidrug resistance-associated protein homolog (MRP4), T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I, in a prostate tissue sample obtained from the subject, wherein an increase in the level of expression of the gene over time is indicative of the progression of the prostate cancer in the tissue.

[0011] Also provided by the invention are methods for monitoring the progression of ovarian cancer in a subject having, or at risk of having, an ovarian cancer. These methods involve measuring a level of expression of at least one gene selected from the group consisting of laminin, alpha 5; vacuolar proton pump, beta polypeptide; putative cytoskeletal protein, natriuretic peptide receptor A, eyes absent homolog, U90916, AL049313, S100 alpha, keratinocyte transglutaminase, GPCR64, meis1, spondin 1, GPCR39, AL050069, mammoglobin 2, and branched chain aminotransferase 1, cytosolic, in an ovarian tissue sample obtained from the subject, wherein an increase in the level of expression of the gene over time is indicative of the progression of the ovarian cancer in the tissue.

[0012] The invention provides methods for identifying agents for use in treatment of prostate cancer comprising. These methods involve:

a) contacting a sample of diseased prostate cells with a candidate agent;
b) detecting a level of expression of at least one gene in the diseased prostate cells, wherein the gene is selected from the group consisting of LIM, multidrug resistance-associated protein homolog (MRP4), T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I; and
c) comparing the level of expression of the gene in the sample in the presence of the candidate agent with a level of expression of the gene in cells that are not contacted with the candidate agent, wherein a decreased level of expression of the gene in the sample in the presence of the candidate agent relative to the expression of the gene in the sample in the absence of the candidate agent is indicative of an agent useful in the treatment of prostate cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] Figure 1. Selection of tumor-specific genes for cancer class prediction. (A) Schematic diagram depicting the idealized expression profile of tumor-specific genes that the method selects as classifiers. The shape of each profile represents genes that are highly expressed in each cancer type relative to all other tumors in the training set.

(B) 100 genes per tumor class (1,100 total) with the most significant scores in a Wilcoxon rank sum test for equality were selected as likely candidates for tumor classifiers. Pr - prostate; Bl - bladder; Br - breast; Co - colorectal; Ga - gastroesophageal; Ki - kidney; Li - liver; Ov - ovary; Pa - pancreas; LA - lung adenocarcinomas; LS - lung squamous cell carcinoma. (C) The final refined set of gene classifiers is generated after ranking genes in (B) by support vector machine (SVM)/ leave-out-one cross-validation (LOOCV) accuracy. Annotations of the genes from which 110 'predictor' genes are bootstrapped are provided in Table 3. For clarity, only 8/76 predictor genes for lung adenocarcinomas are depicted here. Levels of gene expression (depicted in each row) across all samples (columns) are median-centered and normalized by 'Cluster' and output in 'Treeview' (see Eisen et al., *Proc. Nat'l. Acad. Sci. USA*, 95:14863-14868 (1998)). Red - increased gene expression, blue - decreased expression, black - median level of gene expression. The color intensity is proportional to the hybridization intensity of a gene from its median level across all samples.

[0014] Figure 2. Tumor- and tissue-specific genes as class predictors of ovarian and prostate tumors. Shown are the expression levels of highly predictive classifier genes in normal and malignant samples of the ovary and prostate. (A) Expression levels of 28 genes in 5 normal and 24 serous papillary carcinomas of the ovary. (B) Expression levels of 29 genes in 9 normal and 24 localized prostate adenocarcinomas. Genes are conservatively determined to be differentially expressed if the mean level of expression in tumor samples is >3 times the mean level of expression in normal tissues and if p<0.01 (green bars). Gene expression is normalized and output in Treeview as described in Figure 1.

[0015] Figure 3. Detection of the Wilm's Tumor protein (WT) in ovarian cancers. Tissue microarrays containing 36 epithelial tissues and 229 carcinomas representative of the 10 anatomic sites of the tumors profiled in the study are stained with an antibody specific to the WT protein. (A) Visualization of the array using hematoxylin and eosin staining. (B) Normal serous lining of the ovary positive for WT. (C) Three serous papillary carcinomas of the ovary positive for WT. (D) Breast, lung and kidney carcinomas negative for WT (immunoperoxidase technique). Insets show magnified view of nuclei. (E) Transcription of the Wilm's tumor gene (*WT-1*) gene in 175 carcinomas; arrows indicate ovarian tumors in the training and blinded tumor sets. Colors are described in Figure 1.

[0016] Figure 4. Initial analysis of gene expression in the ten most commonly fatal tumors by simple hierarchical clustering.

DETAILED DESCRIPTION

Assays and kits for classifying cell types based on molecular signatures

[0017] This invention provides devices, kits, methods and algorithms for classifying cell types on the basis of their "molecular signatures", such as gene expression profiles. The methods and algorithms are useful, for example, for analyzing the effect of drugs, toxicants or other factors on cells. The present invention relates to the identification of genes that exhibit a characteristic pattern of expression in cells of a particular type or cells that are exposed to a type of stimulus. For example, the invention provides devices and methods by which one can identify the anatomic origin of the ten cancers that are most commonly fatal in the United States (prostate, breast, colorectum, lung, ovary, gastroesophagus, pancreas, liver, kidney and bladder). Subsets of genes whose expression is uniquely characteristic for these carcinomas are identified and used to develop the algorithms and methods of the invention. These algorithms are applied to a mRNA profile or protein expression data from an unknown tumor to determine the type of carcinoma. Such information is key to devising an appropriate treatment strategy. This aspect of the invention is also useful for studies of the mechanistic basis of tumorigenesis, and also finds application in the testing of potential anti-cancer therapeutic agents, and in the diagnosis and prognosis of cancer.

[0018] Thus, in one aspect, the invention provides molecular signatures of the ten most commonly fatal types of cancer. The genes that are expressed in the cancer types include those listed in Table 3. By virtue of their distinctive expression profiles, these genes can be utilized in the diagnosis, management, treatment and/or post-treatment follow-up of persons at risk for, with, or at risk for recurrence of cancers.

[0019] The algorithms and methods of the invention are useful not only for characterizing tumor cells, but also for characterizing other cells that exhibit differential expression of particular genes compared to other cells. For example, a cell that is obtained from an organism that has been exposed to a drug or toxin will generally exhibit differences in expression of one or more genes. By applying the algorithms of the invention, one can determine which of these differences are most predictive of exposure to the drug or toxin. These algorithms for obtaining a molecular signature of gene expression are generally applicable to any cell type.

[0020] Once a molecular signature is obtained, it can be used to analyze cells from a wide variety of samples. For example, a tissue sample can be obtained from the subject, a human or animal model, by known surgical methods, e.g., surgical resection or needle biopsy. A sample of bodily fluid, preferably blood, can also be obtained by standard methods. Plant cells can also be analyzed, as can cells of fungi and microorganisms, including prokaryotes.

[0021] As stated above, the molecular signatures found for particular carcinomas are particularly useful in identifying the anatomic origin, i.e., the tissue origin, of tumors present in a subject. The tissue origin of the tumor found in a subject, e.g., an animal, preferably a human, is of prostate, breast, colorectal, lung, ovarian, gastroesophageal, pancreatic, liver, kidney, or bladder tissue origin.

[0022] In a particularly useful embodiment, a method for identifying a tissue origin of a tumor in a subject comprises:

a) obtaining a sample of the tumor from the subject;
b) detecting a level of expression of at least one gene in each gene set designated for each cancer type as identified in Table 3, in the subject sample to provide a first value; and
c) comparing the first value with a level of expression of the gene in each gene set designated for each cancer type as identified in Table 3, in a sample obtained from a subject of each cancer type, wherein a greater level of expression of the gene in one gene set in the subject sample compared with the level of expression of the gene in each cancer type sample indicates the tissue origin of the tumor.

[0023] The tumor present in the subject can be a metastatic lesion or a primary tumor whose cellular features of tissue origin are not readily identifiable.

[0024] The cancer types identified in Table 3 include prostate (PR), bladder (BL), breast (BR), colorectal (CO), gastroesophageal (GA), kidney (KI), liver (LI), ovary (OV), pancreatic (PA), lung adenoma (LU_A), and lung squamous (LU_S) cancer types. Typically, the level of expression of the gene from the gene set designated for the particular cancer type is about 2-, 5-, 10- or 100- fold or more than the expression level of that gene in the other cancer types.

[0025] A sample of tumor can be taken from the subject by methods well known in the art such as a biopsy. A sample obtained from a subject of each of the cancer types identified in Table 3 can be obtained from different individuals having a specific cancer

type, or can be a pre-established control for which expression of the gene in each gene set selected for each cancer type was determined at an earlier time.

[0026] In some embodiments of this method, it is desirable to determine the level of 2, 3, 5, 10 or more of the genes in each gene set designated for each cancer type as identified in Table 3.

[0027] The level of expression of at least one of the genes that make up the molecular signature in the samples obtained from the subject can be detected by measuring either the level of mRNA corresponding to the gene or the protein encoded by the gene. RNA can be isolated from the samples by methods well-known to those skilled in the art as described, e.g., in Ausubel et al., *Current Protocols in Molecular Biology*, 1:4.1.1-4.2.9 and 4.5.1-4.5.3, John Wiley & Sons, Inc. (1996). Methods for detecting the level of expression of mRNA are well-known in the art and include, but are not limited to, northern blotting, reverse transcription PCR, real time quantitative PCR and other hybridization methods.

[0028] A particularly useful method for detecting the level of mRNA transcripts expressed from a plurality of the disclosed genes involves hybridization of labeled mRNA to an ordered array of oligonucleotides. Such a method allows the level of transcription of a plurality of these genes to be determined simultaneously to generate gene expression profiles or patterns.

[0029] The oligonucleotides utilized in this hybridization method are typically bound to a solid support. Examples of solid supports include, but are not limited to, membranes, filters, slides, paper, nylon, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, polymers, polyvinyl chloride dishes, etc. Any solid surface to which the oligonucleotides can be bound, either directly or indirectly, either covalently or non-covalently, can be used. A particularly preferred solid substrate is a high-density array or DNA chip. These high-density arrays contain a particular oligonucleotide probe in a preselected location on the array. Each preselected location can contain more than one molecule of the particular probe. Because the oligonucleotides are at specified locations on the substrate, the hybridization patterns and intensities (which together result in a unique expression profile or pattern) can be interpreted in terms of expression levels of particular genes.

[0030] The oligonucleotide probes are preferably of sufficient length to specifically hybridize only to complementary transcripts of the above identified gene(s) of

interest. As used herein, the term "oligonucleotide" refers to a single-stranded nucleic acid. Generally the oligonucleotides probes will be at least 16-20 nucleotides in length, although in some cases longer probes of at least 20-25 nucleotides will be desirable.

[0031] Once the probes are contacted with mRNA (or a cDNA copy) obtained from the, the presence of hybridized mRNA or cDNA from the sample is detected by methods known to those of skill in the art. For example, oligonucleotide probes can be labeled with one or more labeling moieties to permit detection of the hybridized probe/target polynucleotide complexes. Label moieties can include compositions that can be detected by spectroscopic, biochemical, photochemical, bioelectronic, immunochemical, electrical optical or chemical means. Examples of labeling moieties include, but are not limited to, radioisotopes, e.g., ^{32}P, ^{33}P, ^{35}S, chemiluminescent compounds, labeled-binding proteins, heavy metal atoms, spectroscopic markers, such as fluorescent markers and dyes, linked enzymes, mass spectrometry tags and magnetic labels.

[0032] Oligonucleotide probe arrays for expression monitoring can be prepared and used according to techniques which are well-known to those skilled in the art as described, e.g., in Lockhart et al., *Nature Biotechnol.*, 14:1675-1680 (1996); McGall et al., *Proc. Nat'l. Acad. Sci. USA*, 93:13555-13460 (1996); and U.S. Patent No. 6,040,138. Such DNA chips are commercially available from, for example, Affymetrix (Santa Clara, CA).

[0033] One can also detect expression of a protein encoded by one or more of the gene(s) that comprise the molecular signature. This can be accomplished by well-established methods, such as, for example, use of a probe that is detectably-labeled, or which can be subsequently-labeled. Generally, the probe is an antibody which recognizes the expressed protein. As used herein, the term antibody includes, but is not limited to, polyclonal antibodies, monoclonal antibodies, humanized or chimeric antibodies and biologically functional antibody fragments which are those fragments sufficient for binding of the antibody fragment to the protein.

[0034] For the production of antibodies to a protein encoded by one of the disclosed genes or to a fragment of the protein, various host animals may be immunized by injection with the polypeptide, or a portion thereof. Such host animals may include, but are not limited to, rabbits, mice and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including, but not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide,

surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (*bacille Calmette-Guerin*) and *Corynebacterium parvum.*

[0035] Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as target gene product, or an antigenic functional derivative thereof. For the production of polyclonal antibodies, host animals, such as those described above, may be immunized by injection with the encoded protein, or a portion thereof, supplemented with adjuvants as also described above.

[0036] Monoclonal antibodies (mAbs), which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler and Milstein (Nature, Vol. 256, pp. 495-497 (1975); and U.S. Patent No. 4,376,110), the human B-cell hybridoma technique (Kosbor et al., Immunology Today, Vol. 4, p. 72 (1983); Cole et al., Proc. Natl. Acad. Sci. USA, Vol. 80, pp. 2026-2030 (1983)), and the EBV-hybridoma technique (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96 (1985)). Such antibodies may be of any immunoglobulin class, including IgG, IgM, IgE, IgA, IgD, and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

[0037] In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., Proc. Natl. Acad. Sci. USA, Vol. 81, pp. 6851-6855 (1984); Neuberger et al., Nature, Vol. 312, pp. 604-608 (1984); Takeda et al., Nature, Vol. 314, pp. 452-454 (1985)) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity, together with genes from a human antibody molecule of appropriate biological activity, can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable or hypervariable region derived from a murine mAb and a human immunoglobulin constant region.

[0038] Alternatively, techniques described for the production of single-chain antibodies (U.S. Patent No. 4,946,778; Bird, Science, Vol. 242, pp. 423-426 (1988); Huston et al., Proc. Natl. Acad. Sci. USA, Vol. 85, pp. 5879-5883 (1988); and Ward et al., Nature,

Vol. 334, pp. 544-546 (1989)) can be adapted to produce differentially expressed gene-single chain antibodies. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single-chain polypeptide.

[0039] Most preferably, techniques useful for the production of "humanized antibodies" can be adapted to produce antibodies to the proteins, fragments or derivatives thereof. Such techniques are disclosed in U.S. Patent Nos. 5,932,448; 5,693,762; 5,693,761; 5,585,089; 5,530,101; 5,569,825; 5,625,126; 5,633,425; 5,789,650; 5,661,016; and 5,770,429.

[0040] Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments, which can be produced by pepsin digestion of the antibody molecule, and the Fab fragments, which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed (Huse et al., Science, Vol. 246, pp. 1275-1281 (1989)) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

[0041] The extent to which the known proteins are expressed in the sample is then determined by immunoassay methods which utilize the antibodies. Such immunoassay methods include, but are not limited to, dot blotting, western blotting, competitive and non-competitive protein binding assays, enzyme-linked immunosorbant assays (ELISA), immunohistochemistry, fluorescence-activated cell sorting (FACS), and others commonly used and widely-described in scientific and patent literature, and many employed commercially.

[0042] A particularly preferred immunoassay method for determining the level of expression of a large number of proteins that make up a molecular signature for a cell type is an antibody array. In this technique, antibodies, preferably monoclonal antibodies specific for the proteins of interest, are directly deposited at high density on a support, e.g., high density array. Similar technology has also been developed for preparing high density DNA microarrays. (See, e.g., Shalon et al., Genome Research, Vol. 6, pp. 639-645 (1996)). The antibody array is then incubated with a protein sample, e.g., a tumor sample from a subject as described above, which is prepared under conditions that reduce native protein-protein interactions. Following incubation, any unbound or non-specific binding proteins can be removed by washing. The proteins that are specifically bound to their respective antibodies

on the array can then be detected. Since the antibodies are bound to the array in a predetermined order, the identity of the protein bound at each position can be ascertained. Measurement of the quantity of protein at all positions on the array thus reflects the protein expression pattern in the sample. The quantity of proteins bound to the array can be measured by several well known methods. For example, the proteins in the sample can be metabolically labeled with radioactive isotopes, e.g., ^{35}S for total proteins and ^{32}P for phosphorylated proteins. The amount of labeled proteins bound to each antibody on the array can be measured by autoradiography and densitometry. The protein sample can also be labeled by biotinylation *in vitro*. The biotinylated proteins bound on the array can then be detected by avidin or streptavidin which binds to biotin. If the avidin is conjugated with horseradish peroxidase or alkaline phosphatase, the bound protein can be visualized by enhanced chemical luminescence. The quantity of protein bound to each antibody indicates the level of that particular protein in the sample. Other methods can also be used to detect the proteins bound to the antibody array, e.g., immunochemical staining and matrix-assisted laser desorption/ionization-time of flight.

[0043] The invention also provides antibody-based panels for identifying the tissue origin, i.e., the anatomic site of origin, of a tumor in a subject. The panel comprises a set of antibody reagents, wherein the set includes at least one antibody reagent specific for detecting a protein encoded by at least one gene in each gene set designated for each cancer type as identified in Table 3. The tissue origin of the tumor is of prostate, breast, colorectal, lung, ovarian, gastroesophageal, pancreatic, liver, kidney, or bladder tissue origin. The term "antibody" is defined above and is preferably a monoclonal antibody specific for detecting the protein.

[0044] In some embodiments of the antibody-based panel, the set includes 2, 3, 5, 10 or more antibody reagents specific for detecting proteins encoded by 2, 3, 5, 10 or more genes, respectively, in each gene set designated for each cancer type as identified in Table 3.

[0045] The invention also provides devices for use in classifying cell types. For example, the invention provides DNA microarrays that include probes for two or more of the genes that make up an expression profile of a particular cell type. In presently preferred embodiments, each array will include probes that are diagnostic for two or more cell types. An array for characterizing cancers could include, for example, probes for some or all of the genes shown in Table 3 that are diagnostic of two or more of the indicated solid tumor types.

[0046] The invention also provides antibody arrays that include antibodies specific for at least one protein encoded by a gene in a gene set designated for each cancer type as identified in Table 3. Preferably the antibody arrays includes antibodies specific for 2, 3, 5, 10 or more proteins encoded by the respective genes in each gene set designated for each cancer type as set forth in Table 3.

[0047] A number of the genes in each gene set that distinguish one tumor type from another as identified in Table 3 are also found to be overexpressed in the different tumor types when compared to normal tissue. These tumor-specific genes can be utilized as biomarkers for the diagnosis, management, treatment and post-treatment of the various cancers described herein. For example, Figures 2A and 2B lists genes (identified by the bar) that are tumor-specific for ovarian cancer and prostate cancer, respectively. Identification of tumor-specific genes from the other gene sets designated for the cancer types including breast, colorectum, lung, ovary, gastroesophagus, pancreas, liver, kidney and bladder, can be readily determined by measuring the level of expression of the genes in a sample obtained from a each cancer type and comparing it to the level of expression of the genes in a sample obtained from the respective normal tissues. An increase in the level of expression of a gene(s) in the set of genes that classify the particular cancer type relative the the level of expression of the gene(s) in its respective normal tissue indicates that the gene(s) is a tumor-specific gene.

[0048] Accordingly, in one aspect, the invention also provides for diagnostic and prognostic assays which are capable of detecting differential expression of specific genes in ovarian and prostate cancers compared with normal ovarian and prostate tissues.

[0049] In one embodiment, a method for screening a subject for prostate cancer or at risk of developing prostate cancer is provided which comprises:

a) detecting a level of expression of at least one gene in a sample of prostate tissue obtained from the subject to provide a first value, wherein the gene is selected from the group consisting of LIM, multidrug resistance-associated protein homolog (MRP4), T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I; and
b) comparing the first value with a level of expression of the gene in a sample of prostate tissue obtained from a disease-free subject, wherein a greater expression level in the subject sample compared to the sample from the disease-free subject is

indicative of the subject having prostate cancer or at risk of developing prostate cancer.

[0050] In another embodiment, a method for screening a subject for ovarian cancer or at risk of developing ovarian cancer is provided which comprises:

a) detecting a level of expression of at least one gene in a sample of ovarian tissue obtained from the subject to provide a first value, wherein the gene is selected from the group consisting of laminin, alpha 5; vacuolar proton pump, beta polypeptide; putative cytoskeletal protein, natriuretic peptide receptor A, eyes absent homolog, U90916, AL049313, S100 alpha, keratinocyte transglutaminase, GPCR64, meis1, spondin 1, GPCR39, AL050069, mammoglobin 2, and branched chain aminotransferase 1, cytosolic; mesothelin and kallikrein 6.

b) comparing the first value with a level of expression of the gene in a sample of ovarian tissue obtained from a disease-free subject, wherein a greater expression level in the subject sample compared to the sample from the disease-free subject is indicative of the subject having ovarian cancer or at risk of developing ovarian cancer.

[0051] The prostate or ovarian tissue sample can be obtained from the subject, a human or animal model, by known surgical methods, e.g., surgical resection or needle biopsy. The sample taken from the disease-free subject can be a sample of normal prostate or ovarian tissue or bodily fluid from the same individual or from another individual. For example, in examination of a suspected prostate or ovarian cancer, the sample from the disease-free subject can be a sample of normal prostate or ovarian cells from the individual suspected of having prostate or ovarian cancer. These normal cells can be obtained from a site adjacent to the tissue suspected of containing the prostate or ovarian cells. Alternatively, the sample taken from the disease-free subject can be a sample of normal prostate or ovarian tissue obtained from another individual. The sample obtained from the disease-free subject can be obtained at the same time as the sample obtained from the subject, or can be a pre-established control for which expression of the gene was determined at an earlier time. The level of expression of the gene in the sample obtained from the disease-free subject is determined and quantitated using the same approach as used for the sample obtained from the subject.

[0052] The level of expression of at least one of the disclosed genes in the samples obtained from the subject and disease-free subject can be detected by measuring either the level of mRNA corresponding to the gene, the protein encoded by the gene or a fragment of the protein by methods well known in the art as described above. In the methods of the invention, the level of expression of one of the disclosed genes in a diseased prostate or ovarian tissue preferably differs from the level of expression of the gene in a non-diseased tissue by a statistically significant amount. In presently preferred embodiments, at least about a 2-fold difference in expression levels is observed. In some embodiments, the expression levels of a gene differ by at least about 5-, 10- or 100-fold or more in the diseased tissue compared to the non-diseased tissue.

[0053] In preferred embodiment of these methods, the level of expression of two, three or more genes is detected.

[0054] The invention also provides for methods of monitoring the progression of a cancer, e.g., a prostate or ovarian cancer, in a subject by measuring a level of expression of mRNA corresponding to, or protein encoded by, at least one of the tumor-specific genes that are differentially expressed in the cancer, in a sample obtained from the subject over time, i.e., at various stages of the disease. An increase in the level of expression of the gene(s) over time is indicative of the progression of the cancer. The level of expression of the gene(s) can be detected by standard methods as described above.

<u>Assays to identify agents that modulate expression</u>

[0055] In another aspect, a cell-based assay based on one or more of the genes that make up a molecular signature can be used to identify agents that modify the expression of these genes. Such agents find use, for example, in the treatment of the condition (e.g., a particular type of cancer) for which the molecular signature is diagnostic. These methods typically involve: a) contacting a sample obtained from a subject suspected of having the condition of interest with a candidate agent; b) detecting a level of expression of at least one gene that comprises the molecular signature (e.g., for cancer, a gene identified in Table 3); and c) comparing the level of expression of the gene in the sample in the presence of the candidate agent with a level of expression of the gene in the sample in the absence of the candidate agent, wherein an increased or decreased level of expression in the sample in the presence of the agent relative to the level of expression in the absence of the agent is

indicative of an agent that can modulate the expression of the gene. The level of expression of the gene can be detected by, for example, measuring the level of mRNA corresponding to or protein encoded by the gene as described above. In presently preferred embodiments, the expression of more than one gene in the molecular signature is monitored for modulation by the candidate agent.

[0056] As used herein, the term "candidate agent" refers to any molecule that is capable of decreasing the level of mRNA corresponding to or protein encoded by at least one of the genes that comprise a molecular signature. The candidate agents can be natural or synthetic molecules such as proteins or fragments thereof, small molecule inhibitors, nucleic acid molecules, e.g., antisense nucleotides, ribozymes, double-stranded RNAs, organic and inorganic compounds and the like.

[0057] In particular, the cell-based assay can be utilized to identify agents that inhibit or decrease the expression of one or more genes that are differentially expressed, i.e., overexpressed in diseased cells, i.e., cancer cells, compared to non-diseased cells. As stated above, genes that are overexpressed in cancer tissue relative to the respective normal tissue can be discerned by measuring the expression level of the gene in a sample of both tissues and comparing the expression levels obtained for both tissues. Figures 2A and 2B disclose tumor-specific genes that are overexpressed in ovarian and prostate cancers, respectivel. Other cancer cells include breast, colorectal, gastroesophageal, pancreatic, liver, kidney and bladder cells.

[0058] In one embodiment, a method for identifying agents for use in treatment of prostate cancer is provided which comprises:

a) contacting a sample of diseased prostate cells with a candidate agent;
b) detecting a level of expression of at least one gene in the diseased prostate cells, wherein the gene is selected from the group consisting of LIM, multidrug resistance-associated protein homolog (MRP4), T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I; and

c) comparing the level of expression of the gene in the sample in the presence of the candidate agent with a level of expression of the gene in cells that are not contacted with the candidate agent, wherein a decreased level of expression of the gene in the sample in the presence of the candidate agent relative to the expression of the gene in the sample in the absence of the candidate agent is indicative of an agent useful in the treatment of prostate

cancer.

[0059] In another embodiment, a method of identifying agents useful in the treatment of ovarian cancer is provided which comprises:

a) contacting a sample of diseased ovarian cells with a candidate agent;

b) detecting a level of expression of at least one gene in the diseased ovarian cells, wherein the gene is selected from the group consisting of laminin, alpha 5; vacuolar proton pump, beta polypeptide; putative cytoskeletal protein, natriuretic peptide receptor A, eyes absent homolog, U90916, AL049313, S100 alpha, keratinocyte transglutaminase, GPCR64, meis1, spondin 1, GPCR39, AL050069, mammoglobin 2, and branched chain aminotransferase 1, cytosolic; and

c) comparing the level of expression of the gene in the sample in the presence of the candidate agent with a level of expression of the gene in cells that are not contacted with the candidate agent, wherein a decreased level of expression of the gene in the sample in the presence of the candidate agent relative to the expression of the gene in the sample in the absence of the candidate agent is indicative of an agent useful in the treatment of ovarian cancer.

[0060] Cell-free assays can also be used to identify compounds which are capable of interacting with a protein encoded by one or more of the genes that make up the molecular signature, or with a binding partner of one of these encoded proteins, to alter the activity of the protein or its binding partner. Cell-free assays can also be used to identify compounds which modulate the interaction between the encoded protein and its binding partner, such as a target peptide. In one embodiment, cell-free assays for identifying such compounds comprise a reaction mixture containing a protein encoded by one of the molecular signature component genes and a test compound or a library of test compounds in the presence or absence of the binding partner, e.g., a biologically inactive target peptide, or a small molecule. Accordingly, one example of a cell-free method for identifying agents useful in the modulation of the underlying condition for which the molecular signature is characteristic involves contacting a protein or functional fragment thereof or the protein binding partner with a test compound or library of test compounds and detecting the formation of complexes. For detection purposes, the protein can be labeled with a specific marker and the test compound or library of test compounds labeled with a different marker. Interaction of a test compound with the protein or fragment thereof or the protein binding

partner can then be detected by measuring the level of the two labels after incubation and washing steps. The presence of the two labels is indicative of an interaction.

[0061] Interaction between molecules can also be assessed by using real-time BIA (Biomolecular Interaction Analysis, Pharmacia Biosensor AB) which detects surface plasmon resonance, an optical phenomenon. Detection depends on changes in the mass concentration of mass macromolecules at the biospecific interface and does not require labeling of the molecules. In one useful embodiment, a library of test compounds can be immobilized on a sensor surface, e.g., a wall of a micro-flow cell. A solution containing the protein, functional fragment thereof, or the protein binding partner is then continuously circulated over the sensor surface. An alteration in the resonance angle as indicated on a signal recording, indicates the occurrence of an interaction. This technique is described in more detail in BIA Technology Handbook by Pharmacia.

[0062] Another embodiment of a cell-free assay involves: a) combining a protein encoded by the gene, the protein binding partner, and a test compound to form a reaction mixture; and b) detecting interaction of the protein and the protein binding partner in the presence and absence of the test compounds. A considerable change (potentiation or inhibition) in the interaction of the protein and binding partner in the presence of the test compound compared to the interaction in the absence of the test compound indicates a potential agonist (mimetic or potentiator) or antagonist (inhibitor) of the protein's activity for the test compound. The components of the assay can be combined simultaneously or the protein can be contacted with the test compound for a period of time, followed by the addition of the binding partner to the reaction mixture. The efficacy of the compound can be assessed by using various concentrations of the compound to generate dose response curves. A control assay can also be performed by quantitating the formation of the complex between the protein and its binding partner in the absence of the test compound.

[0063] Formation of a complex between the protein and its binding partner can be detected by using detectably-labeled proteins, such as radiolabeled, fluorescently-labeled or enzymatically-labeled protein or its binding partner, by immunoassay or by chromatographic detection.

[0064] In preferred embodiments, the protein or its binding partner can be immobilized to facilitate separation of complexes from uncomplexed forms of the protein and its binding partner and automation of the assay. Complexation of the protein to its

binding partner can be achieved in any type of vessel, e.g., microtitre plates, microcentrifuge tubes and test tubes. In particularly preferred embodiment, the protein can be fused to another protein, e.g., glutathione-S-transferase to form a fusion protein which can be adsorbed onto a matrix, e.g., glutathione Sepharose® beads (Sigma Chemical, St. Louis, MO) which are then combined with the labeled protein partner, e.g., labeled with ^{35}S, and test compound and incubated under conditions sufficient to formation of complexes. Subsequently, the beads are washed to remove unbound label and the matrix is immobilized and the radiolabel is determined.

[0065] Another method for immobilizing proteins on matrices involves utilizing biotin and streptavidin. For example, the protein can be biotinylated using biotin N-hydroxy-succinimide (NHS) using well-known techniques and immobilized in the well of steptavidin-coated plates.

[0066] Cell-free assays can also be used to identify agents which are capable of interacting with a protein encoded by at least one gene that comprises a molecular signature and modulate the activity of the protein encoded by the gene. In one embodiment, the protein is incubated with a test compound and the catalytic activity of the protein is determined. In another embodiment, the binding affinity of the protein to a target molecule can be determined by methods known in the art.

[0067] The present invention also provides for both prophylactic and therapeutic methods of treating a subject having or at risk of having a disorder or condition for which a molecular signature is diagnostic. Subjects at risk for such disorders can be identified by a prognostic assay, e.g., as described above. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the disorder or condition, such that development of the disorder is prevented or delayed in its progression. With respect to treatment of the disorder, it is not required that the cell, e.g., cancer cell, be killed or induced to undergo cell death. Instead, all that is required to achieve treatment of the disorder is that the tumor growth be slowed down to some degree or that some of the abnormal cells revert back to normal. Examples of suitable therapeutic agents include, but are not limited to, antisense nucleotides, ribozymes, double-stranded RNAs and antagonists. The molecular signatures of the invention are useful for monitoring the efficacy of a particular course of treatment for the disorder or condition.

[0068] As used herein, the term "antisense" refers to nucleotide sequences that are complementary to a portion of an RNA expression product of at least one of the disclosed genes. "Complementary" nucleotide sequences refer to nucleotide sequences that are capable of base-pairing according to the standard Watson-Crick complementarity rules. That is, purines will base- pair with pyrimidine to form combinations of guanine:cytosine and adenine:thymine in the case of DNA, or adenine:uracil in the case of RNA. Other less common bases, e.g., inosine, 5-methylcytosine, 6-methyladenine, hypoxanthine and others may be included in the hybridizing sequences and will not interfere with pairing.

[0069] When introduced into a host cell, antisense nucleotide sequences specifically hybridize with the cellular mRNA and/or genomic DNA corresponding to the gene(s) so as to inhibit expression of the encoded protein, e.g., by inhibiting transcription and/or translation within the cell.

[0070] The isolated nucleic acid molecule comprising the antisense nucleotide sequence can be delivered, e.g., as an expression vector, which when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the encoded mRNA of the gene(s). Alternatively, the isolated nucleic acid molecule comprising the antisense nucleotide sequence is an oligonucleotide probe which is prepared *ex vivo* and, which, when introduced into the cell, results in inhibiting expression of the encoded protein by hybridizing with the mRNA and/or genomic sequences of the gene(s).

[0071] Preferably, the oligonucleotide contains artificial internucleotide linkages which render the antisense molecule resistant to exonucleases and endonucleases, and thus are stable in the cell. Examples of modified nucleic acid molecules for use as antisense nucleotide sequences are phosphoramidate, phosporothioate and methylphosphonate analogs of DNA as described, e.g., in U.S. Patent No. 5,176,996; 5,264,564; and 5,256,775. General approaches to preparing oligomers useful in antisense therapy are described, e.g., in Van der Krol, BioTechniques, Vol. 6, pp. 958-976 (1988); and Stein et al., Cancer Res., Vol. 48, pp. 2659-2668 (1988).

[0072] Typical antisense approaches, involve the preparation of oligonucleotides, either DNA or RNA, that are complementary to the encoded mRNA of the gene. The antisense oligonucleotides will hybridize to the encoded mRNA of the gene and prevent translation. The capacity of the antisense nucleotide sequence to hybridize with the desired gene will depend on the degree of complementarity and the length of the antisense

nucleotide sequence. Typically, as the length of the hybridizing nucleic acid increases, the more base mismatches with an RNA it may contain and still form a stable duplex or triplex. One skilled in the art can determine a tolerable degree of mismatch by use of conventional procedures to determine the melting point of the hybridized complexes.

[0073] Antisense oligonucleotides are preferably designed to be complementary to the 5' end of the mRNA, e.g., the 5' untranslated sequence up to and including the regions complementary to the mRNA initiation site, i.e., AUG. However, oligonucleotide sequences that are complementary to the 3' untranslated sequence of mRNA have also been shown to be effective at inhibiting translation of mRNAs as described, e.g., in Wagner, Nature, Vol. 372, pp. 333 (1994). While antisense oligonucleotides can be designed to be complementary to the mRNA coding regions, such oligonucleotides are less efficient inhibitors of translation.

[0074] Regardless of the mRNA region to which they hybridize, antisense oligonucleotides are generally from about 15 to about 25 nucleotides in length.

[0075] The antisense nucleotide can also comprise at least one modified base moiety, e.g., 3-methylcytosine, 5,-methylcytosine, 7-methylguanine, 5-fluorouracil, 5-bromouracil, and may also comprise at least one modified sugar moiety, e.g., arabinose, hexose, 2-fluorarabinose, and xylulose.

[0076] In another embodiment, the antisense nucleotide sequence is an alpha-anomeric nucleotide sequence. An alpha-anomeric nucleotide sequence forms specific double stranded hybrids with complementary RNA, in which, contrary to the usual beta-units, the strands run parallel to each other as described e.g., in Gautier et al., Nucl. Acids. Res., Vol. 15, pp. 6625-6641 (1987).

[0077] Antisense nucleotides can be delivered to cells which express the described genes *in vivo* by various techniques, e.g., injection directly into the prostate tissue site, entrapping the antisense nucleotide in a liposome, by administering modified antisense nucleotides which are targeted to the prostate cells by linking the antisense nucleotides to peptides or antibodies that specifically bind receptors or antigens expressed on the cell surface.

[0078] However, with the above-mentioned delivery methods, it may be difficult to attain intracellular concentrations sufficient to inhibit translation of endogenous mRNA. Accordingly, in a preferred embodiment, the nucleic acid comprising an antisense nucleotide

sequence is placed under the transcriptional control of a promoter, i.e., a DNA sequence which is required to initiate transcription of the specific genes, to form an expression construct. The use of such a construct to transfect cells results in the transcription of sufficient amounts of single stranded RNAs to hybridize with the endogenous mRNAs of the described genes, thereby inhibiting translation of the encoded mRNA of the gene. For example, a vector can be introduced *in vivo* such that it is taken up by a cell and directs the transcription of the antisense nucleotide sequence. Such vectors can be constructed by standard recombinant technology methods. Typical expression vectors include bacterial plasmids or phage, such as those of the pUC or Bluescript.TM plasmid series, or viral vectors such as adenovirus, adeno-associated virus, herpes virus, vaccinia virus and retrovirus adapted for use in eukaryotic cells. Expression of the antisense nucleotide sequence can be achieved by any promoter known in the art to act in mammalian cells. Examples of such promoters include, but are not limited to, the promoter contained in the 3' long terminal repeat of Rous sarcoma virus as described, e.g., in Yamamoto et al., Cell, Vol. 22, pp. 787-797 (1980); the herpes thymidine kinase promoter as described, e.g., in Wagner et al., Proc. Natl. Acad. Sci. USA, Vol. 78, pp. 1441-1445 (1981); the SV40 early promoter region as described, e.g., in Bernoist and Chambon, Nature, Vol. 290, pp. 304-310 (1981); and the regulatory sequences of the metallothionein gene as described, e.g., in Brinster et al., Nature, Vol. 296, pp. 39-42 (1982).

[0079] Ribozymes are RNA molecules that specifically cleave other single-stranded RNA in a manner similar to DNA restriction endonucleases. By modifying the nucleotide sequences encoding the RNAs, ribozymes can be synthesized to recognize specific nucleotide sequences in a molecule and cleave it as described, e.g., in Cech, J. Amer. Med. Assn., Vol. 260, p. 3030 (1988). Accordingly, only mRNAs with specific sequences are cleaved and inactivated.

[0080] Two basic types of ribozymes include the "hammerhead"-type as described for example in Rossie et al., Pharmac. Ther., Vol. 50, pp. 245-254 (1991); and the hairpin ribozyme as described, e.g., in Hampel et al., Nucl. Acids Res., Vol. 18, pp. 299-304 (1999) and U.S. Patent No. 5,254,678. Intracellular expression of hammerhead and hairpin ribozymes targeted to mRNA corresponding to at least one of the disclosed genes can be utilized to inhibit protein encoded by the gene.

[0081] Ribozymes can either be delivered directly to cells, in the form of RNA oligonucleotides incorporating ribozyme sequences, or introduced into the cell as an expression vector encoding the desired ribozymal RNA. Ribozyme sequences can be modified in essentially the same manner as described for antisense nucleotides, e.g., the ribozyme sequence can comprise a modified base moiety.

[0082] Double-stranded RNA, i.e., sense-antisense RNA, corresponding to at least one of the disclosed genes, can also be utilized to interfere with expression of at least one of the disclosed genes. Interference with the function and expression of endogenous genes by double-stranded RNA has been shown in various organisms such as *C. elegans* as described, e.g., in Fire et al., Nature, Vol. 391, pp. 806-811 (1998); drosophilia as described, e.g., in Kennerdell et al., Cell, Vol. 95, No. 7, pp. 1017-1026 (1998); and mouse embryos as described, e.g., in Wianni et al., Nat. Cell Biol., Vol. 2, No. 2, pp. 70-75 (2000). Such double-stranded RNA can be synthesized by *in vitro* transcription of single-stranded RNA read from both directions of a template and *in vitro* annealing of sense and antisense RNA strands. Double-stranded RNA can also be synthesized from a cDNA vector construct in which the gene of interest is cloned in opposing orientations separated by an inverted repeat. Following cell transfection, the RNA is transcribed and the complementary strands reanneal. Double-stranded RNA corresponding to at least one of the disclosed genes could be introduced into a prostate cell by cell transfection of a construct such as that described above.

[0083] The term "antagonist" refers to a molecule which, when bound to the protein encoded by the gene, inhibits its activity. Antagonists can include, but are not limited to, peptides, proteins, carbohydrates, and small molecules.

[0084] In a particularly useful embodiment, the antagonist is an antibody specific for the protein expressed by the gene. Antibodies useful as therapeutics encompass the antibodies as described above. The antibody alone may act as an effector of therapy or it may recruit other cells to actually effect cell killing. The antibody may also be conjugated to a reagent such as a chemotherapeutic, radionuclide, ricin A chain, cholera toxin, pertussis toxin, etc., and serve as a target agent. Alternatively, the effector may be a lymphocyte carrying a surface molecule that interacts, either directly or indirectly, with a tumor target. Various effector cells include cytotoxic T cells and NK cells.

[0085] Examples of the antibody-therapeutic agent conjugates which can be used in therapy include, but are not limited to: 1) antibodies coupled to radionuclides, such as ^{125}I, ^{131}I, ^{123}I, ^{111}In, ^{105}Rh, ^{153}Sm, ^{67}Cu, ^{67}Ga, ^{166}Ho, ^{177}Lu, ^{186}Re and ^{188}Re, and as described, e.g., in Goldenberg et al., Cancer Res., Vol. 41, pp. 4354-4360 (1981); Carrasquillo et al., Cancer Treat. Rep., Vol. 68, pp. 317-328 (1984); Zalcberg et al., J. Natl. Cancer Inst., Vol. 72, pp. 697-704 (1984); Jones et al., Int. J. Cancer, Vol. 35, pp. 715-720 (1985); Lange et al., Surgery, Vol. 98, pp. 143-150 (1985); Kaltovich et al., J. Nucl. Med., Vol. 27, p. 897 (1986); Order et al., Int. J. Radiother. Oncol. Biol. Phys., Vol. 8, pp. 259-261 (1982); Courtenay-Luck et al., Lancet, Vol. 1, pp. 1441-1443 (1984); and Ettinger et al., Cancer Treat. Rep., Vol. 66, pp. 289-297 (1982); (2) antibodies coupled to drugs or biological response modifiers such as methotrexate, adriamycin, and lymphokines such as interferon as described, for, e.g., in Chabner et al., Cancer, Principles and Practice of Oncology, Philadelphia, Pa., J. B. Lippincott Co. Vol. 1, pp. 290-328 (1985); Oldham et al., Cancer, Principles and Practice of Oncology, Philadelphia, Pa., J. B. Lippincott Co., Vol. 2, pp. 2223-2245 (1985); Deguchi et al., Cancer Res., Vol. 46, pp. 3751-3755 (1986); Deguchi et al., Fed. Proc., Vol. 44, p. 1684 (1985); Embleton et al., Br. J. Cancer, Vol. 49, pp. 559-565 (1984); and Pimm et al., Cancer Immunol. Immunother., Vol. 12, pp. 125-134 (1982); (3) antibodies coupled to toxins, as described, for example, in Uhr et al., Monoclonal Antibodies and Cancer, Academic Press, Inc., pp. 85-98 (1983); Vitetta et al., Biotechnology and Bio. Frontiers, Ed. P. H. Abelson, pp. 73-85 (1984); and Vitetta et al., Science, Vol. 219, pp. 644-650 (1983); (4) heterofunctional antibodies, for example, antibodies coupled or combined with another antibody so that the complex binds both to the carcinoma and effector cells, e.g., killer cells such as T cells, as described, for example, in Perez et al., J. Exper. Med., Vol. 163, pp. 166-178 (1986); and Lau et al., Proc. Natl. Acad. Sci. USA, Vol. 82, pp. 8648-8652 (1985); and (5) native, i.e., non-conjugated or non-complexed antibodies, as described in, for example, Herlyn et al., Proc. Natl. Acad. Sci. USA, Vol. 79, pp. 4761-4765 (1982); Schulz et al., Proc. Natl. Acad. Sci. USA, Vol. 80, pp. 5407-5411 (1983); Capone et al., Proc. Natl. Acad. Sci. USA, Vol. 80, pp. 7328-7332 (1983); Sears et al., Cancer Res., Vol. 45, pp. 5910-5913 (1985); Nepom et al., Proc. Natl. Acad. Sci. USA, Vol. 81, pp. 2864-2867 (1984); Koprowski et al., Proc. Natl. Acad. Sci. USA, Vol. 81, pp. 216-219 (1984); and Houghton et al., Proc. Natl. Acad. Sci. USA, Vol. 82, pp. 1242-1246 (1985).

[0086] Methods for coupling an antibody or fragment thereof to a therapeutic agent as described above are well known in the art and are described, e.g., in the methods provided in the references above. In yet another embodiment, the antagonist useful as a therapeutic for treating cancer, e.g., prostate cancer, can be an inhibitor of a protein encoded by one of the disclosed genes.

[0087] In the case of treatment with an antisense nucleotide, the method comprises administering a therapeutically effective amount of an isolated nucleic acid molecule comprising an antisense nucleotide sequence derived from at least one gene identified in Table 3, Figure 2A or Figure 2B. In one embodiment, the gene is preferably an ovarian tumor-specific gene selected from the group consisting of laminin, alpha 5; vacuolar proton pump, beta polypeptide; putative cytoskeletal protein, natriuretic peptide receptor A, eyes absent homolog, U90916, AL049313, S100 alpha, keratinocyte transglutaminase, GPCR64, meis1, spondin 1, GPCR39, AL050069, mammoglobin 2, and branched chain aminotransferase 1, cytosolic 4 as is described in Figure 2A.

[0088] In another embodiment, the gene is preferably a prostate tumor-specific gene selected from the group consisting of LIM, multidrug resistance-associated protein homolog (MRP4); T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I as is disclosed in Figure 2B. The term "isolated" nucleic acid molecule means that the nucleic acid molecule is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally occurring nucleic acid molecule is not isolated, but the same nucleic acid molecule, separated from some or all of the co-existing materials in the natural system, is isolated, even if subsequently reintroduced into the natural system. Such nucleic acid molecules could be part of a vector or part of a composition and still be isolated, in that such vector or composition is not part of its natural environment.

[0089] With respect to treatment with a ribozyme or double-stranded RNA molecule, the method comprises administering a therapeutically effective amount of a nucleotide sequence encoding a ribozyme, or a double-stranded RNA molecule, wherein the nucleotide sequence encoding the ribozyme/double-stranded RNA molecule has the ability to decrease the transcription/translation of at least one gene identified in Table 3, Figures 2A or 2B, and is preferably the ovarian and tumor-specific genes disclosed in Figures 2A and 2B, respectively.

[0090] In the case of treatment with an antagonist, the method comprises administering to a subject a therapeutically effective amount of an antagonist that inhibits a protein encoded by at least one gene identified in Table 3, Figure 2A or Figure 2B, and is preferably the ovarian and prostate tumor-specific genes disclosed in Figures 2A or B, respectively.

[0091] A "therapeutically effective amount" of an isolated nucleic acid molecule comprising an antisense nucleotide, nucleotide sequence encoding a ribozyme, double-stranded RNA, or antagonist, refers to a sufficient amount of one of these therapeutic agents to treat a cancer, e.g., a prostate cancer (e.g., to limit prostate tumor growth or to slow or block tumor metastasis). The determination of a therapeutically effective amount is well within the capability of those skilled in the art. For any therapeutic, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

[0092] Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Antisense nucleotides, ribozymes, double-stranded RNAs, and antagonists which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range, depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

[0093] The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the

subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy.

[0094] Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for antagonists.

[0095] For therapeutic applications, the antisense nucleotides, nucleotide sequences encoding ribozymes, double-stranded RNAs (whether entrapped in a liposome or contained in a viral vector) and antagonists are preferably administered as pharmaceutical compositions containing the therapeutic agent in combination with one or more pharmaceutically acceptable carriers. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones.

[0096] The pharmaceutical compositions may be administered by any number of routes, including, but not limited to, oral, intravenous, intramuscular, intra-articular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

[0097] In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.).

[0098] Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

[0099] Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

[0100] Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

[0101] Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

[0102] Pharmaceutical formulations suitable for parenteral administration may be formulated from aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain

suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

[0103] For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

[0104] The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

[0105] The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0. 1-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

[0106] After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of the antisense nucleotide or antagonist, such labeling would include amount, frequency, and method of administration. Those skilled in the art will employ different formulations for antisense nucleotides than for antagonists, e.g., antibodies or inhibitors. Pharmaceutical formulations suitable for oral administration of proteins are described, e.g., in U.S. Patent Nos. 5,008,114; 5,505,962; 5,641,515; 5,681,811; 5,700,486; 5,766,633; 5,792,451; 5,853,748; 5,972,387; 5,976,569; and 6,051,561.

[0107] In another aspect, the treatment of a subject with a therapeutic agent, such as those described above, can be monitored by detecting the level of expression of mRNA or protein encoded by at least one of the disclosed genes identified in Table 3, Figure 2A or Figure 2B. These measurements will indicate whether the treatment is effective or whether it should be adjusted or optimized. Accordingly, one or more of the genes described herein can be used as a marker for the efficacy of a drug during clinical trials.

[0108] In a particularly useful embodiment, a method for monitoring the efficacy of a treatment of a subject having a prostate or ovarian cancer. or at risk of. or having such a

cancer with an agent (e.g., an antagonist, protein, nucleic acid, small molecule, or other therapeutic agent or candidate agent identified by the screening assays described herein) is provided comprising:

a) obtaining a pre-administration sample from a subject prior to administration of the agent,

b) detecting the level of expression of mRNA corresponding to, or protein encoded by the gene, or activity of the protein encoded by the gene identified in Table 3, Figure 2A or Figure 2B in the pre-administration sample;

c) obtaining one or more post-administration samples from the subject,

d) detecting the level of expression of mRNA corresponding to, or protein encoded by the gene, or activity of the protein encoded by the gene in the post-administration sample or samples,

e) comparing the level of expression of mRNA or protein encoded by the gene, or activity of the protein encoded by the gene in the pre-administration sample with the level of expression of mRNA or protein encoded by the gene, or activity of the protein encoded by the gene in the post-administration sample or samples, and

f) adjusting the administration of the agent accordingly.

[0109] For example, increased administration of the agent may be desirable to decrease the level of expression or activity of the gene to lower levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to increase expression or activity of the gene to higher levels than detected, i.e., to decrease the effectiveness of the agent.

[0110] In another aspect, a method for inhibiting undesired proliferation of a cancer cell, particularly a prostate or ovarian cell is provided which utilizes a therapeutic agent as described above, e.g., an antisense nucleotide, a ribozyme, a double-stranded RNA, and an antagonist such as an antibody. Preferably, the prostate or ovarian cell is present in a human. The undesired proliferation of the prostate or ovarian cell is associated with a condition including, but not limited to localized prostate cancer, metastatic prostate cancer, benign prostatic hyperplasia and ovarian cancer.

[0111] With respect to inhibition of proliferation of a prostate or ovarian cell utilizing an antisense nucleotide, the method comprises administering to the prostate or ovarian cell a therapeutically effective amount of an isolated nucleic acid molecule comprising an antisense nucleotide sequence derived from at least one gene identified in Tables 3, Figure 2A or Figure 2B, wherein the antisense nucleotide has the ability to decrease the transcription/translation of the gene.

[0112] With respect to inhibition of proliferation of a prostate or ovarian cell utilizing a ribozyme, such a method comprises administering to the prostate or ovarian cell a therapeutically effective amount of a nucleotide sequence encoding the ribozyme, which has the ability to decrease the transcription/translation of at least one gene identified in Table 3, Figure 2A or Figure 2B.

[0113] With respect to inhibition of proliferation of a prostate or ovarian cell utilizing a double-stranded RNA, the method comprises administering to the prostate cell a therapeutically effective amount of a double-stranded RNA corresponding to at least one gene identified in Table 3, Figure 2A or Figure 2B.

[0114] With respect to inhibition of proliferation of a prostate cell utilizing an antagonist, the method comprises administering to the prostate or ovarian cell a therapeutically effective amount of an antagonist that inhibits a protein encoded by at least one gene identified in Table 3, Figure 2A or Figure 2B.

[0115] In the context of inhibiting undesired proliferation of a cancer cell such as prostate or ovarian cell, a "therapeutically effective amount" of an isolated nucleic acid molecule comprising an antisense nucleotide, a nucleotide sequence encoding a ribozyme, a double-stranded RNA, or antagonist, refers to a sufficient amount of one of these therapeutic agents to inhibit proliferation of a cancer cell (e.g., to inhibit or stabilize cellular growth of the cancer cell) and can be determined as described above.

[0116] In another aspect, a viral vector is provided which comprises a promoter and/or an enhancer or other regulatory element of a gene selected from the group consisting of at least one of the genes identified in Table 3, Figures 2A or 2B, and preferably the tumor-specific genes disclosed for Figures 2 or 3, operably linked to the coding region of a gene that is essential for replication of the vector, wherein the vector is adapted to replicate upon transfection into a diseased prostate cell. The promoter sequences can be discerned by searching the publicly available databases for BAC clones that cover the entire gene;

thereafter, the cDNA for the gene can be compared to the genomic sequence. This will generally reveal the intron-exon boundaries and the start site of the gene. Once these are established, the promoter sequences can be inferred. Such vectors are able to selectively replicate in a cancer cell such as prostate cell, but not in a non-diseased cancer cell. The replication is conditional upon the presence in a diseased cancer cell, and not in a non-diseased cancer cell, of positive transcription factors that activate the promoter of the disclosed genes selected for each cancer, e.g., prostate cancer as identified in Table 3 or the prostate tumor-specific genes disclosed in Figure 2B. It can also occur by the absence of transcription inhibiting factors that normally occur in a non-diseased cell, e.g., a prostate cell, and prevent transcription as a result of the promoter. Accordingly, when transcription occurs, it proceeds into the gene essential for replication, such that in the diseased cell, but not in non-diseased cell, replication of the vector and its attendant functions occur. With this vector, a diseased prostate cell, e.g., a prostate cancer cell, can be selectively treated, with minimal systemic toxicity.

[0117] In one embodiment, the viral vector is an adenoviral vector, which includes a coding region of a gene essential for replication of the vector, wherein the coding region is selected from the group consisting of E1a, E1b, E2 and E4 coding regions. The term "gene essential for replication" refers to a nucleic acid sequence whose transcription is required for the vector to replicate in the target cell. Preferably, the gene essential for replication is selected from the group consisting of the E1A and E1b coding sequences. Particularly preferred is the adenoviral E1A gene as the gene essential for replication. Methods for making such vectors are well know to the person of ordinary skill in the art as described, e.g., in Sambrook et al., in Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, New York, 1989. The present invention provides novel viral vectors based on the oncolytic adenoviral vector strategy as described in U.S. Patent No. 5,998,205, issued December 7, 1999 to Hallenbeck et al. and in U.S. provisional application ____________ filed January 14, 2002, entitled "Novel Oncolytic Adenoviral Vectors" (Docket No. 4-31704P3/PROV/GTI), the disclosures of which are hereby incorporated by reference in their entirety. In particular, oncolytic adenoviral vectors are disclosed in which expression of at least one adenoviral gene, which is essential for replication, is controlled by a tissue-specific promoter which is selectively transactivated in cancer cells. In one embodiment a tissue-specific promoter controls the expression of E1a. In a particularly preferred embodiment

both the E1a and E4 genes are controlled by tumor-specific promoters. Methods for preparing tissue-specific replication vectors and their use in the treatment of cancer cells and other types of abnormal cells which are harmful or otherwise unwanted *in vivo* in a subject are described in detail, e.g., in U.S. Patent No. 5,998,205. U.S. Patent No. 5,698,443 describes adenoviral vectors, in which expression of a gene essential for replication is controlled by the PSA promoter/enhancer. Unlike the vectors of the present invention, however, the viral vectors described in this patent replicate in normal as well as diseased prostate cells, because PSA promoter/enhancer is active in normal cells as well as in diseased cells.

[0118] In a further embodiment, the invention provides nucleic acid constructs in which a heterologous gene product is expressed under the control of a promoter and/or an enhancer or other regulatory element of a gene selected from the group consisting of at least one of the genes identified in Table 3, Figures 2A and 2B, and is preferably selected from the tumor-specific genes disclosed in Figures 2A and 2B. Such heterologous gene products are expressed when the construct is present in diseased cells, e.g., cancer cells, but not in normal, non-diseased cells. The heterologous gene product provides, in some embodiments, for the inhibition, prevention, or destruction of the growth of the diseased cell, e.g., a prostate cancer cell. The gene product can be RNA, e.g., antisense RNA or ribozyme, or proteins such as a cytokine, e.g., interleukin, interferon, or toxins such as diphtheria toxin, pseudomonas toxin, etc. The heterologous gene product can also be a negative selective marker such as cytosine deaminase. Such negative selective markers can interact with other agents to prevent, inhibit or destroy the growth of the diseased prostate or ovarian cells. U.S. Patent No. 6,057,299, for example, describes the construction and use of nucleic acid constructs in which heterologous genes are placed under the control of a PSA enhancer. The nucleic acid constructs can be introduced into target cells by methods known to those of skill in the art. For example, one can incorporate the constructs into an appropriate vector such as those described above.

[0119] The vector of the present invention can be transfected into a helper cell line for viral replication and to generate infectious viral particles. Alternatively, transfection of the vector or other nucleic acid into a cancer cell can take place by electroporation, calcium phosphate precipitation, microinjection, or through liposomes, including proteoliposomes.

EXAMPLE

[0120] The following example is offered to illustrate, but not to limit the present invention.

[0121] This Example describes the use of mRNA profiling of the ten most commonly fatal carcinomas, coupled with supervised machine learning algorithms, to identify subsets of genes whose expression is uniquely characteristic for each of these ten carcinomas. These genes were used to accurately predict the anatomic origin of 75 blinded carcinomas, including metastatic lesions, with up to 95% success rates. This study demonstrates the existence of subsets of genes whose transcription is characteristic of specific carcinomas, despite a wide-ranging appearance of the tumor cells, and illustrates the feasibility of predicting the anatomic site of tumor origin in the context of multiple diverse tumor classes.

[0122] A global approach to this problem is taken by identifying sets of genes whose expression is specific to carcinomas of the prostate, breast, colorectum, lung, ovary, gastroesophagus, pancreas, liver, kidney and bladder, which together account for ~70% (~400,000 cases) of all cancer-related deaths in the United States (see Greenlee et al., *CA Cancer J. Clin.*, 50:7-33 (2000)). mRNA from 100 carefully dissected primary tumors is analyzed with oligonucleotide microarrays containing detectors for 12,533 genes to obtain quantitative measurements of gene transcription in each sample. The initial set of 100 primary carcinomas is comprised of 10 prostate adenocarcinomas, 9 bladder carcinomas (8 transitional cell carcinomas and 1 squamous cell carcinoma), 10 infilitrating ductal breast carcinomas, 10 colorectal adenocarcinomas, 11 gastroesophageal adenocarcinomas, 11 kidney carcinomas, 6 liver (hepatocellular) carcinomas, 10 serous papillary ovarian adenocarcinomas, 6 pancreatic carcinomas and 17 lung carcinomas (9 adenocarcinomas and 8 squamous cell carcinomas). Each specimen is assessed by frozen section examination, and areas rich in tumor are cut from the frozen blocks prior to RNA extraction. Care is taken to avoid non-neoplastic epithelium within the tumor samples. RNA extraction and hybridization is performed as described (see Lockhart et al., *Nature Biotechnol.*, 14:1675-1680 (1996); and Wodicka et al., *Nature Biotechnol.*, 15:1359-1367 (1997)), with the exception that the arrays are hybridized at 50°C for 16-20 hours. GeneChip® hybridization data are processed and scaled as described (see Lockhart, *supra*; and Wodicka, *supra*)). Only those probe sets (9,198) are included whose maximum hybridization intensity (average

difference (AD)) across all samples is >200; the other probe sets are excluded. All AD values <20, including negative AD values, are raised to a value of 20 and the data is log transformed. A complete description of all of the tumors used in this study and the primary hybridization data are available from our website (www.gnf.org/cancer/epican/tumors). Cancer classification schemes are then developed to identify specific sets of genes that could be used as classifiers to predict the anatomic origin of 75 unknown tumor samples. This provides a quantitative measure of the extent to which these genes are characteristic of an individual tumor type. Finally, individual genes are further characterized in the classifier sets to determine tissue versus tumor specific expression.

[0123] Initial analysis of gene expression in these tumors by simple hierarchical clustering reveals complex patterns of transcription. Although separate cancers of some anatomic sites, such as the prostate and kidney, can be readily separated based solely on the patterns of the most variably expressed genes, a striking degree of similarity is identified between cancers of the colorectum, stomach, bladder and lung (see Figure 4), making histologic separation difficult. Therefore the process of multi-class prediction is divided into three components: a) filtering the large data set of gene expression (12,533 genes in 100 tumors, >1.25 million data-points) to exclude those genes that do not contribute to tumor distinction; b) ranking potentially predictive genes to identify the most accurate tumor-specific classifiers; and c) determining an optimal method by which these genes could be used to 'vote' for the likely class of a blinded tumor sample in the context of multiple tumor classes.

[0124] Genes in each tumor class that have the most significant probabilities of being differentially expressed relative to all other classes are first identified by a Wilcoxon rank test (see Figure 1). For each of the 9,198 genes, a Wilcoxon rank score is calculated for the group with the highest mean expression *versus* samples from all other groups (implemented in Matlab v6.0). One hundred genes from each tumor class identified by this procedure are then subjected to a 'prediction accuracy test', in which each gene is individually evaluated for its ability to discriminate one tumor class from all other tumor classes using a supervised machine learning classifier (see Figure 1). The 100 genes with the lowest p-values in each class (total 1,100 genes) are ranked based on their predictive accuracy for discriminating one tumor class *versus* all others using a support vector machine (SVM) classifier. Specifically, genes are ranked based on their LOOCV accuracy. In

LOOCV for a given gene, we blind ourselves to one sample, trained an SVM using the remaining samples, and use the SVM to predict the class identity of the blinded sample (either cancer class X, or not cancer class X). This process is repeated for all samples in the training set, and an overall prediction accuracy is calculated for each gene. The SVM procedure (E. Dimitriadou, K. Hornik, F. Leisch, D. Meyer, and A. Weingessel) is implemented in the software package R v1.2.2.4.

[0125] A voting scheme is developed based on calculating a 'class distance', by which to evaluate how molecularly related an unknown sample is to tumors of different classes. The voting scheme utilizes the 10 genes with the highest SVM/LOOCV accuracy from each class (110 total genes). For each class, a minimum SVM/LOOCV accuracy threshold is set such that at least ten genes passed; since in each class multiple genes have equivalent accuracy, 216 genes are selected from the 11 classes and were iteratively bootstrapped to obtain an equal number (10) of voting genes per class. For classifying an unknown sample, prediction scores are calculated using one set of 110 genes (calculated as described below), and final predictions are based on averaged scores over 50 iterations. Hybridization values for our 110-gene predictor set are compared to each sample in our training set. An L1 distance (sum of absolute differences) from the unknown sample to each training sample is calculated. The "class distance" is defined as the mean distance from the unknown sample to the members of that class in the training set. The class to which an unknown sample has the lowest class distance is the predicted identity.

[0126] A confidence score is also employed to estimate the strength of each prediction, and experimentally determines a confidence threshold that minimizes tumor misclassification. A Dixon test for outliers is employed to assign a confidence score to each prediction. The Dixon metric is calculated by sorting the vector of mean distances, where $x_i < x_{i+1}$, and computing the value $D = (x_2 - x_1) / (x_n - x_1)$. A Dixon threshold of $D = 0.1$ is empirically set as a conservative boundary for high confidence predictions. Empirically, it is determined that a small group of 110 genes, representing 10 genes per tumor class, most accurately predicts the origin of a blinded tumor sample (see Table 3).

[0127] Using these optimized parameters, the performance of the classification method is first assessed by predicting an anatomic site of origin for each of 100 tumors in the training set by cross-validation (see Tables 1 and 4). Confident predictions are made for

94/100 (94%) of the samples, of which 92 (98%) are correct. The 6 unclassified cases do not pass the confidence threshold imposed on the experiment.

[0128] The classification scheme is then applied to an independent series of 75 cancer samples, which are blinded during training of the classifier. This group is comprised of tumors with histologies represented in our training set, including 12 metastatic lesions and many poorly differentiated tumors whose cellular features are not entirely indicative of their anatomic origin. Specifically, the set of 75 blinded tumor samples includes 63 primary tumors and 12 metastatic lesions. The primary tumor samples are 9 lung cancers (4 adenocarcinomas, 5 squamous cell carcinoma, 9 colorectal adenocarcinomas, 13 breast carcinomas, 14 prostate adenocarcinomas, 15 papillary serous ovarian carcinomas, 1 hepatocellular carcinoma and 2 gastroesophageal carcinomas). More detailed description of the ovarian and prostate cancer collection has been reported (see Welsh et al., *Proc. Nat'l. Acad. Sci. USA*, 98:1176-1181 (2002); and Welsh et al., *Submitted for publication* (2001 Confident and accurate predictions for 64/75 (85%) tumors are made above the empirically set confidence threshold, including 9/12 (75%) metastatic cancers. None of these tumors samples are incorrectly classified. In the absence of the confidence threshold, an anatomic origin of 97/100 (97%) tumors is correctly predicted in the training set by cross-validation and 71/75 (95%) tumors in the blinded sample set, including 11/12 (92%) metastatic lesions (see Tables 1 and 4).

Table 1. Prediction accuracy based on a 100-tumor training set

Tumor Set	Number of tumors	Dixon Confidence Threshold			No Confidence Threshold	
		Correct	Misclassified	No Call	Correct	Misclassified
Training set (cross-validation)	100	92 (92%)	2 (2%)	6 (6%)	97 (97%)	3 (3%)
Blinded set	75	64 (85%)	0 (0%)	11 (15%)	71 (95%)	4 (5%)

Two different groups of tumors were predicted using our classification method: 100 tumors comprising the training set (Training set) and a group of 75 tumors (Blinded set). Each sample in the training set was blinded and predicted in a cross-validation study. The blinded set contained samples not included in the training set, the identities of which were unknown during the training and optimization of the method.

[0129] Classification of tumors arising in certain anatomic sites is relatively straightforward because of the large number of unequivocal predictor genes (e.g., 19 genes with 100% predictive accuracy for prostate cancer). In contrast, prediction of other tumors,

such as those of the lung, bladder or gastroesophagus, is more difficult because of the relative paucity of highly predictive classifier genes. The difficulty in selecting genes whose expression is specific to these cancers reflects a high degree of molecular relatedness, which we have observed upon initial analyses of tumor gene expression (see Figure 4). For example, blinded gastroesophageal cancers that could not be predicted by our method are assigned as lung tumors (albeit with confidence scores close to zero; see Tables 2 and 4). Analysis of the entire human transcriptome with these and other tumors may identify additional tumor-specific genes that would augment those identified here.

Table 2. Distribution of class predictions

		Predicted Class										
		PR	BL	BR	CO	GA	KI	LI	OV	PA	LU_A	LU_S
True identity of unknown sample	PR	26 (0.564)										
	BL		8 (0.343)									
	BR			26 (0.267)								
	CO				23 (0.279)							
	GA					11 (0.187)					2 (0.044)	
	KI						11 (0.502)					
	LI			1 (0.115)				5 (0.523)			1 (0.041)	
	OV		1 (0.045)						26 (0.317)			
	PA									6 (0.529)		
	LA		1 (0.020)								13 (0.275)	
	LS			1 (0.138)								13 (0.307)

[0130] The genes that constitute the most accurate cancer class predictors (see Table 3) are divisible into two groups: those whose altered transcription is characteristic of specific neoplasms (termed 'tumor-specific'), and those that are characteristic of the tissue in which they are normally expressed, rather than of the cancers that arise in these tissues (termed 'tissue-specific'). On the basis of gene annotation alone, many well-described tumor-specific markers and targets are recognized. These include *MUC-2* and *A33* in colon

cancers, the latter of which has been used as an immunotherapeutic target in advanced colorectal carcinomas (see Tschmelitsch et al., *Cancer Res.*, 57:2181-2186 (1997)); mammaglobin-1 (*MGB-1*) and uroplakin II (*UPII*), which have been proposed as highly sensitive diagnostic markers for micro-metastatic breast and bladder cancers, respectively (see Ghossein et al., *In vivo*, 14:237-250 (2000); and Li et al., *J. Urol.*, 162:931-935 (1999)); and thyroid transcription factor 1 (*TFF-1*), which has been proposed as a highly accurate marker for differential diagnosis of lung adenocarcinomas (see Reis-Filho et al., *Pathol. Res. Pract.*, 196:835-840 (2000)). Examples of tissue-specific genes are kidney organic cation transporter, liver serum albumin and pancreatic lipase (see Table 3).

[0131] Interestingly, genes are also identified whose annotations suggest their expression in the stromal cells that surround epithelial tumors. In some cases, evidence is subsequently found of their over-expression in malignant epithelia (e.g., the fibroblast activation protein (*FAP-α*) in breast cancers (see Kelly et al., *Mod. Pathol.*, 11:855-863 (1998)). In adenocarcinomas of the lung, genes are identified whose annotations indicate the presence of B-cells, T-cells, macrophages and neutrophils, reflecting a positive smoking history in these patients. With the exception of *TTF-1*, few of the genes identified in lung adenocarcinoma (see Table 3) are good classifiers.

[0132] Because of the inherent difficultly in using gene annotation alone to judge tumor-specific *versus* tissue-specific gene expression, it is sought to objectively 'dissect' predictor gene subsets into these different components. The levels of expression of 28 and 29 highly-ranked predictor genes for ovarian and prostate cancers, respectively, are directly compared in normal and tumor tissues. The 29 prostate cancer-specific genes are chosen that are all >99% predictive of prostate cancer within the training set of 101 tumors, and 28 ovarian cancer-specific genes, which are at least 92% predictive of ovarian cancers. The expression levels of these genes in an expanded set of 24 ovarian and 24 prostate cancer samples are compared against 5 and 9 normal samples of ovary and prostate, respectively (see Welsh, *supra*; and Welsh, *supra*). Genes are ranked by an unpaired *t*-test and a measure of mean difference in expression levels. Differential expression is determined for genes whose expression is significantly different in normal and tumor tissues ($p<0.01$) and where the mean level of expression in tumor tissues is >3 times that in normal tissues. In ovarian cancers, 18/28 genes are significantly over-expressed in the tumors (see Figure 2A). Among this group of genes are protease M/neurosin/kallikrein 6 (*hK6*), which is a candidate serum

marker for ovarian cancer (see Diamandis et al., *Clin. Biochem.*, 33:579-583 (2000)), and mesothelin (*CAK1*), which is over-expressed in ovarian cancers and used as a specific target for a novel therapeutic immunotoxin (see Hassan et al., *J. Immunother.*, 20:2902-2906 (2000)). Two G-protein coupled receptors, *GPR39* and *GPR64*, are also identified which are important examples of potential tumor-specific therapeutic targets discovered by this approach. The 10 tissue-specific genes, which include the *WT-1*, *smad6* and *Hox5.1*, most likely represent features of normal ovarian physiology. In prostate, significant cancer-specific up-regulation of 6 genes is observed (see Figure 2B), including the prostate-specific T-cell receptor gamma chain (*TCRγ*). Two products of the *TCRγ* locus are transcribed, of which the smaller transcript, translated as a 7kDa protein termed *TARP*, is only expressed in the nucleus of prostate and breast cancer cells and is thought to be under the control of estrogens and androgens (see Wolfgang et al., *Proc. Nat'l. Acad. Sci. USA*, 15:9437-9442 (2000)). Over-expression of the calmodulin-dependent kinase 1 (*CAMK1*), testican, the multi-drug resistance gene, *MRP4* (*ABCC4*) and a LIM domain protein is also found. To our knowledge, none of these have yet been reported as over-expressed in prostate cancers.

[0133] To evaluate whether the expression of these genes is tumor-specific in the context of gene expression in a whole array of human tissues, the expression of ovarian, prostate and other tumor-specific genes is analyzed in an expanded set of 46 normal human body tissues organs and cell lines. In ovarian cancers, for example, very few or no body tissues exhibit discernable expression of several of the ovarian cancer predictor genes, including *CAK1* and *hK6*. Importantly, some of the genes with unknown function in several cancer-types are also highly tissue-specific, highlighting the potential of this method to identify novel, highly restricted tumor-specific genes for molecular intervention or diagnosis.

[0134] The increased tumor-specific transcription of a subset of these genes is further evaluated by analysis of over-expression of their protein products. For example, a polyclonal antibody specific to the WT protein is used, whose transcript is highly-expressed in ovarian cancers, on tissue microarrays containing 229 carcinomas representing tumors from the 10 anatomic sites analyzed in the study. The tissue microarrays contain 0.6 mm cores from 265 different zinc formalin-fixed paraffin-embedded specimens and are constructed using a Tissue Microarrayer (Beecher Instruments, Silver Springs, MD). Samples consist of 36 normal adult epithelial tissues and 229 carcinomas that include most

of the tumors whose transcripts are profiled in the study. Ovarian cancers are profiled as previously described (see Welsh, *supra*) and 16 other independent serous papillary carcinomas of the ovary are included in the tissue microarrays. For immunohistochemistry on the tissue microarrays and on a whole tissue section of a normal ovary, the avidin-biotin immunoperoxidase method is performed. After slides had been placed in a citrate buffer and treated with microwave heat for 20 minutes, the polyclonal anti-WT antibody (C-19;1:100 dilution; Santa Cruz Biotechnology, Santa Cruz, CA) is applied for one hour at room temperature. Nuclear immunoreactivity is considered to represent true positivity. Immunostaining for WT protein is present in nuclei from 18/20 (90%) serous papillary carcinomas, while nuclear immunoreactivity is absent in the other 209 carcinomas (see Figure 3). As expected from the analysis of tissue- and tumor-specific transcription in ovarian cancers, the normal serous lining epithelium of the ovary is also positive for WT protein (see Figure 3). The results indicate that the derivation of antibodies to the products of tumor-specific genes described here can have clinical utility for early detection, predicting tumor origin, monitoring patients for tumor recurrence, and possibly antibody-based therapy. This potential is underscored by the identification of many known or predicted tumor diagnostic genes, such as *MGB-1* in breast cancer, *PSA* and *hK2* in prostate cancer, *hK6* in ovarian cancer and uroplakins 1b and II in bladder carcinomas (see Table 3).

[0135] A striking conclusion from the data presented here is that subsets of genes with highly restricted, tumor-specific expression can be identified for as many as 11 distinct tumor classes, despite well-described tumor heterogeneity and obvious molecular similarities among many divergent tumor classes (see Figure 4). The fact, that these gene subsets can be successfully used to predict the origin of a given tumor in a majority of cases, underscores how strongly characteristic these genes are for specific histopathological subtypes of cancer. In that regard, it is worth noting that using as few as 11 genes (i.e., one gene per tumor class), the anatomic origin can be predicted of up to 91% and 83% of the training and blinded tumor samples, respectively. These success rates indicate the broad applicability of these methods to other molecular classification problems, including identifying the molecular signatures of diverse toxicants or drug responses. The groups of predictor genes that are 'tumor-specific' (see Figure 2), including some of those that are discussed above, are especially attractive because class distinction among tumors of diverse tissue origin inherently selects for tissue-specificity. These features are highly relevant for

immunotherapeutic and chemical antagonism of gene function, as well as providing novel targets for gene therapy approaches for selective tumor cell destruction. Finally, these results demonstrate that one can construct custom DNA microarrays for a molecular classification of solid tumors, a resource that will augment traditional site-specific and histopathological classification schemes. This is particularly valuable for cases of metastases where the primary tumor origin is undetermined, estimated at 4% of all patients diagnosed with cancer (see Hillen, *supra*). The extension of these and other discriminant methods to identify molecular correlates with tumor grade, stage, response to therapy and outcome, further contribute to the optimal management of patients with cancer.

Table 3. Annotations of predictor genes

Cancer Type	Affy Probe ID	Name	Description	LOOCV Accuracy
PR	251_at	CAMK1	calcium/calmodulin-dependent protein kinase I	100
PR	37812_at	KIAA0293	KIAA0293 protein	100
PR	40794_at	KLK3	kallikrein 3, (prostate specific antigen)	100
PR	41721_at	KLK2	kallikrein 2, prostatic	100
PR	41468_at	TRG@	T cell receptor gamma locus	100
PR	1513_at	LBX1	transcription factor similar to D. melanogaster homeodomain protein lady bird late	100
PR	41172_at	LOC51109	CGI-82 protein	100
PR	32200_at	ACPP	acid phosphatase, prostate	100
PR	217_at	KLK2	kallikrein 2, prostatic	100
PR	1514_g_at	none	Antigen |TIGR==HG2261-HT2352	100
PR	40297_at	STEAP	six transmembrane epithelial antigen of the prostate	100
PR	617_at	ACPP	acid phosphatase, prostate	100
PR	35778_at	KIF5C	kinesin family member 5C	100
PR	1662_r_at	none	Antigen |TIGR==HG2261-HT2351	100
PR	263_g_at	AMD1	S-adenosylmethionine decarboxylase 1	100
PR	36685_at	AMD1	S-adenosylmethionine decarboxylase 1	100
PR	1661_i_at	none	Antigen |TIGR==HG2261-HT2351	100
PR	40060_r_at	LIM	LIM protein (similar to rat protein kinase C-binding enigma)	100
PR	1805_g_at	KLK3	kallikrein 3, (prostate specific antigen)	100
BL	32448_at	UPK2	uroplakin 2	96
BL	36628_at	RALBP1	ralA binding protein 1	96
BL	36555_at	SNCG	synuclein, gamma (breast cancer-specific protein 1)	96
BL	36571_at	TOP2B	topoisomerase (DNA) II beta (180kD)	96
BL	38457_at	TNNI2	troponin I, skeletal, fast	95
BL	1490_at	MYCL1	v-myc avian myelocytomatosis viral oncogene homolog 1, lung carcinoma derived	95
BL	37104_at	PPARG	peroxisome proliferative activated receptor, gamma	95
BL	39939_at	COL4A6	collagen, type IV, alpha 6	95
BL	32527_at	APM2	adipose specific 2	95
BL	35402_at	DR6	death receptor 6	95
BL	41111_at	BCAT2	branched chain aminotransferase 2, mitochondrial	95
BL	32382_at	UPK1B	uroplakin 1B	95
BR	41348_at	IRX5	iroquois homeobox protein 5	96
BR	33848_r_at	CDKN1B	cyclin-dependent kinase inhibitor 1B (p27, Kip1)	95
BR	33878_at	FLJ13612	hypothetical protein FLJ13612	93
BR	36329_at	MGB1	mammaglobin 1	92

Cancer Type	Affy Probe ID	Name	Description	LOOCV Accuracy
BR	34778_at	none	Homo sapiens cDNA FLJ12280 fis, clone MAMMA1001744	92
BR	37142_at	GFRA1	GDNF family receptor alpha 1	92
BR	686_s_at	none	Homo sapiens endogenous retrovirus HERV-K104 long terminal repeat, complete sequence; and Gag protein (gag) and envelope protein (env) genes, complete cds	92
BR	39945_at	FAP	fibroblast activation protein, alpha	92
BR	40161_at	COMP	cartilage oligomeric matrix protein (pseudoachondroplasia, epiphyseal dysplasia 1, multiple)	92
BR	1177_at	FLJ12443	hypothetical protein FLJ12443	91
BR	40046_r_at	C18ORF1	chromosome 18 open reading frame 1	91
BR	40162_s_at	COMP	cartilage oligomeric matrix protein (pseudoachondroplasia, epiphyseal dysplasia 1, multiple)	91
CO	170_at	CDX2	caudal type homeo box transcription factor 2	97
CO	169_at	CDX1	caudal type homeo box transcription factor 1	95
CO	37423_at	SLC12A2	solute carrier family 12 (sodium/potassium/chloride transporters), member 2	95
CO	38884_at	CLCA1	chloride channel, calcium activated, family member 1	94
CO	37875_at	GPA33	glycoprotein A33 (transmembrane)	94
CO	32972_at	NOX1	NADPH oxidase 1	94
CO	896_at	MUC2	mucin 2, intestinal/tracheal	94
CO	40736_at	CDH17	cadherin 17, LI cadherin (liver-intestine)	93
CO	41073_at	GPR49	G protein-coupled receptor 49	93
CO	37415_at	ATP10B	ATPase, Class V, type 10B	93
CO	41728_at	KIAA0152	KIAA0152 gene product	93
CO	1582_at	CEACAM5	carcinoembryonic antigen-related cell adhesion molecule 5	93
CO	38739_at	ETS2	v-ets avian erythroblastosis virus E26 oncogene homolog 2	93
GA	40957_at	KIAA0160	KIAA0160 protein	96
GA	302_at	AHCYL1	S-adenosylhomocysteine hydrolase-like 1	95
GA	31574_i_at	none	Human HL14 gene encoding beta-galactoside-binding lectin, 3' end, clone 2	94
GA	710_at	P4HB	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase; thyroid hormone binding protein p55)	93
GA	34595_at	MYHL	myosin, heavy polypeptide-like (110kD)	93
GA	38116_at	KIAA0101	KIAA0101 gene product	93
GA	36015_at	CORT	cortistatin	93

Cancer Type	Affy Probe ID	Name	Description	LOOCV Accuracy
GA	31575_f_at	none	Human HL14 gene encoding beta-galactoside-binding lectin, 3' end, clone 2	93
GA	39253_s_at	RALA	v-ral simian leukemia viral oncogene homolog A (ras related)	92
GA	34851_at	STK6	serine/threonine kinase 6	92
GA	40451_at	none	AL080203:Homo sapiens mRNA; cDNA DKFZp434F222 (from clone DKFZp434F222) /cds=(0 |GenBank==AL080203	92
GA	34491_at	OASL	2'-5'-oligoadenylate synthetase-like	92
KI	33534_at	ESM1	endothelial cell-specific molecule 1	100
KI	35220_at	ENPEP	glutamyl aminopeptidase (aminopeptidase A)	100
KI	35243_at	PCTK3	PCTAIRE protein kinase 3	100
KI	39654_at	ASPA	aspartoacylase (aminoacylase 2, Canavan disease)	99
KI	35867_at	SLC22A2	solute carrier family 22 (organic cation transporter), member 2	99
KI	1951_at	ANGPT2	angiopoietin 2	99
KI	39260_at	SLC16A4	solute carrier family 16 (monocarboxylic acid transporters), member 4	99
KI	34777_at	ADM	adrenomedullin	99
KI	40954_at	FXYD2	FXYD domain-containing ion transport regulator 2	99
KI	36668_at	DIA1	diaphorase (NADH) (cytochrome b-5 reductase)	99
LI	37816_at	C5	complement component 5	100
LI	32771_at	GC	group-specific component (vitamin D binding protein)	99
LI	37175_at	SERPINC1	serine (or cysteine) proteinase inhibitor, clade C (antithrombin), member 1	99
LI	37235_g_at	KNG	kininogen	99
LI	37236_at	none	M11437:Human kininogen gene /cds=(0 |GenBank==M11437	99
LI	36342_r_at	HFL3	H factor (complement)-like 3	99
LI	36657_at	APOC2	apolipoprotein C-II	99
LI	40311_at	TFR2	transferrin receptor 2	99
LI	33701_at	PAH	phenylalanine hydroxylase	99
LI	36341_s_at	HFL3	H factor (complement)-like 3	99
LI	39763_at	HPX	hemopexin	99
LI	37202_at	F2	coagulation factor II (thrombin)	99
LI	261_s_at	APOB	apolipoprotein B (including Ag(x) antigen)	99
LI	33377_at	VTN	vitronectin (serum spreading factor, somatomedin B, complement S-protein)	99
LI	36584_at	ITIH2	inter-alpha (globulin) inhibitor, H2 polypeptide	99
LI	35332_at	APOB	apolipoprotein B (including Ag(x) antigen)	99
LI	1232_s_at	IGFBP1	insulin-like growth factor binding protein 1	99

Cancer Type	Affy Probe ID	Name	Description	LOOCV Accuracy
OV	32625_at	NPR1	natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)	99
OV	1500_at	WT1	Wilms tumor 1	99
OV	38201_at	BCAT1	branched chain aminotransferase 1, cytosolic	98
OV	40763_at	MEIS1	Meis1 (mouse) homolog	98
OV	35277_at	SPON1	spondin 1, (f-spondin) extracellular matrix protein	98
OV	40401_at	LOC55816	hypothetical protein	97
OV	34194_at	none	Homo sapiens mRNA; cDNA DKFZp564B076 (from clone DKFZp564B076)	97
OV	32959_at	none	M25809:Human endomembrane proton pump subunit mRNA |GenBank==M25809	96
OV	37554_at	KLK6	kallikrein 6 (neurosin, zyme)	96
OV	35226_at	EYA2	eyes absent (Drosophila) homolog 2	95
OV	38749_at	GPR39	G protein-coupled receptor 39	95
OV	32838_at	none	S67247:smooth muscle myosin heavy chain isoform SMemb [human |GenBank==S67247	95
OV	1955_s_at	MADH6	MAD (mothers against decapentaplegic, Drosophila) homolog 6	95
PA	39177_r_at	CEL	carboxyl ester lipase (bile salt-stimulated lipase)	100
PA	41238_s_at	ELA3	elastase 3, pancreatic (protease E)	100
PA	35594_at	PNLIPRP2	pancreatic lipase-related protein 2	100
PA	31482_at	CELL	carboxyl ester lipase-like (bile salt-stimulated lipase-like)	99
PA	386_g_at	CTRL	chymotrypsin-like	99
PA	36141_at	CTRB1	chymotrypsinogen B1	99
PA	39176_f_at	CEL	carboxyl ester lipase (bile salt-stimulated lipase)	99
PA	39726_at	GCG	glucagon	99
PA	34941_at	CLPS	colipase, pancreatic	99
PA	40714_at	CTRC	chymotrypsin C (caldecrin)	99
PA	31483_g_at	CEL	carboxyl ester lipase (bile salt-stimulated lipase)	99
PA	40043_at	PRSS3	protease, serine, 3 (trypsin 3)	99
PA	40748_at	CPA2	carboxypeptidase A2 (pancreatic)	99
PA	912_s_at	PLA2G1B	phospholipase A2, group IB (pancreas)	99
PA	32796_f_at	PRSS2	protease, serine, 2 (trypsin 2)	99
PA	41369_at	PNLIP	pancreatic lipase	99
PA	34309_at	CPA1	carboxypeptidase A1 (pancreatic)	99
PA	38936_at	ELA1	elastase 1, pancreatic	99
PA	40204_at	CTRL	chymotrypsin-like	99
LU_A	33754_at	TITF1	thyroid transcription factor 1	98
LU_A	40928_at	DKFZP564A122	DKFZP564A122 protein	94
LU_A	40749_at	MS4A2	membrane-spanning 4-domains, subfamily A, member 2 (Fc fragment of IgE, high affinity I, receptor for; beta polypeptide)	94
LU_A	130_s_at	TITF1	thyroid transcription factor 1	94

Cancer Type	Affy Probe ID	Name	Description	LOOCV Accuracy
LU_A	37634_at	PAEP	progestagen-associated endometrial protein (placental protein 14, pregnancy-associated endometrial alpha-2-globulin, alpha uterine protein)	93
LU_A	34876_at	none	U65090:Human carboxypeptidase D mRNA |GenBank==U65090	93
LU_A	33383_f_at	ASAHL	N-acylsphingosine amidohydrolase (acid ceramidase)-like	93
LU_A	32116_at	LAK-4P	expressed in activated T/LAK lymphocytes	92
LU_A	31901_at	KCNAB2	potassium voltage-gated channel, shaker-related subfamily, beta member 2	91
LU_A	32640_at	ICAM1	intercellular adhesion molecule 1 (CD54), human rhinovirus receptor	91
LU_A	40019_at	EVI2B	ecotropic viral integration site 2B	91
LU_A	37148_at	LILRB3	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 3	91
LU_A	31457_at	FOXD2	forkhead box D2	91
LU_A	38332_at	MGC11256	hypothetical protein MGC11256	91
LU_A	40520_g_at	PTPRC	protein tyrosine phosphatase, receptor type, C	91
LU_A	32793_at	TRB@	T cell receptor beta locus	91
LU_A	41165_g_at	none	X67301:H.sapiens mRNA for IgM heavy chain constant region (Ab63) /cds=(0 |GenBank==X67301	91
LU_A	1478_at	ITK	IL2-inducible T-cell kinase	91
LU_A	41164_at	none	X67301:H.sapiens mRNA for IgM heavy chain constant region (Ab63) /cds=(0 |GenBank==X67301	91
LU_A	33956_at	MD-2	MD-2 protein	91
LU_A	32616_at	LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	91
LU_A	37402_at	RNASE1	ribonuclease, RNase A family, 1 (pancreatic)	91
LU_A	36878_f_at	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	91
LU_A	38894_g_at	NCF4	neutrophil cytosolic factor 4 (40kD)	91
LU_A	40518_at	PTPRC	protein tyrosine phosphatase, receptor type, C	91
LU_A	32632_g_at	GBA	glucosidase, beta; acid (includes glucosylceramidase)	91
LU_A	36773_f_at	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	91
LU_A	37351_at	UP	uridine phosphorylase	91
LU_A	31680_at	TOP1	topoisomerase (DNA) I	91
LU_A	34874_at	NTE	neuropathy target esterase	91
LU_A	37637_at	RGS3	regulator of G-protein signalling 3	91
LU_A	37344_at	HLA-DMA	major histocompatibility complex, class II, DM alpha	91

Cancer Type	Affy Probe ID	Name	Description	LOOCV Accuracy
LU_A	1633_g_at	PIM2	pim-2 oncogene	91
LU_A	1461_at	NFKBIA	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	91
LU_A	402_s_at	ICAM3	intercellular adhesion molecule 3	91
LU_A	37218_at	BTG3	BTG family, member 3	91
LU_A	195_s_at	CASP4	caspase 4, apoptosis-related cysteine protease	91
LU_A	37170_at	none	AB015331:Homo sapiens HRIHFB2017 mRNA |GenBank==AB015331	91
LU_A	37411_at	ACAP1	KIAA0050 gene product	91
LU_A	40476_s_at	ILF1	interleukin enhancer binding factor 1	91
LU_A	36440_at	none	Human pre TCR alpha mRNA, partial cds	91
LU_A	1062_g_at	IL10RA	interleukin 10 receptor, alpha	91
LU_A	32558_at	PIAS3	protein inhibitor of activated STAT3	91
LU_A	32715_at	VAMP8	vesicle-associated membrane protein 8 (endobrevin)	91
LU_A	1173_g_at	none	Spermidine/Spermine N1-Acetyltransferase |TIGR==HG172-HT3924	91
LU_A	40406_at	MST1	macrophage stimulating 1 (hepatocyte growth factor-like)	91
LU_A	1665_s_at	none	Homo sapiens CDA02 mRNA, complete cds	91
LU_A	32193_at	none	Homo sapiens clone 23785 mRNA sequence	91
LU_A	39134_at	TOM1	target of myb1 (chicken) homolog	91
LU_A	35132_at	MYO1E	myosin IE	91
LU_A	39296_at	FBLN1	fibulin 1	91
LU_A	568_at	PRKACA	protein kinase, cAMP-dependent, catalytic, alpha	91
LU_A	38378_at	CD53	CD53 antigen	91
LU_A	33273_f_at	IGLJ3	immunoglobulin lambda joining 3	91
LU_A	31526_f_at	USP6	ubiquitin specific protease 6 (Tre-2 oncogene)	91
LU_A	33274_f_at	IGLJ3	immunoglobulin lambda joining 3	91
LU_A	37864_s_at	none	Y14737:Homo sapiens mRNA for immunoglobulin lambda heavy chain /cds=(65 |GenBank==Y14737	91
LU_A	33354_at	SMURF2	E3 ubiquitin ligase SMURF2	91
LU_A	36493_at	LSP1	lymphocyte-specific protein 1	91
LU_A	1427_g_at	SLA	Src-like-adapter	91
LU_A	32091_at	KIAA0446	KIAA0446 gene product	91
LU_A	34677_f_at	none	Homo sapiens mRNA for TL132	91
LU_A	39649_at	ARHGAP4	Rho GTPase activating protein 4	91
LU_A	31820_at	HCLS1	hematopoietic cell-specific Lyn substrate 1	91
LU_A	927_s_at	MUC1	mucin 1, transmembrane	91
LU_A	41827_f_at	IGLL1	immunoglobulin lambda-like polypeptide 1	91
LU_A	41091_at	FALZ	fetal Alzheimer antigen	91
LU_A	38762_at	RNAHP	RNA helicase-related protein	91

Cancer Type	Affy Probe ID	Name	Description	LOOCV Accuracy
LU_A	1729_at	TRADD	TNFRSF1A-associated via death domain	91
LU_A	38869_at	KIAA1069	KIAA1069 protein	91
LU_A	37352_at	SP100	nuclear antigen Sp100	91
LU_A	35248_at	SLC19A2	solute carrier family 19 (thiamine transporter), member 2	91
LU_A	40635_at	FLOT1	flotillin 1	91
LU_A	37975_at	CYBB	cytochrome b-245, beta polypeptide (chronic granulomatous disease)	91
LU_A	34304_s_at	SAT	spermidine/spermine N1-acetyltransferase	91
LU_A	38666_at	PSCD1	pleckstrin homology, Sec7 and coiled/coil domains 1(cytohesin 1)	91
LU_S	39015_f_at	KRT6A	keratin 6A	98
LU_S	39016_r_at	KRT6A	keratin 6A	97
LU_S	33529_at	ADH7	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide	96
LU_S	1560_g_at	PAK2	p21 (CDKN1A)-activated kinase 2	96
LU_S	33693_at	DSG3	desmoglein 3 (pemphigus vulgaris antigen)	96
LU_S	31791_at	TP63	tumor protein 63 kDa with strong homology to p53	96
LU_S	32563_at	ATP1B3	ATPase, Na+/K+ transporting, beta 3 polypeptide	96
LU_S	33108_i_at	SOX2	SRY (sex determining region Y)-box 2	95
LU_S	32380_at	PKP1	plakophilin 1 (ectodermal dysplasia/skin fragility syndrome)	95
LU_S	1932_at	ABCC5	ATP-binding cassette, sub-family C (CFTR/MRP), member 5	95
LU_S	36457_at	GMPS	guanine monophosphate synthetase	95
LU_S	39581_at	CSTA	cystatin A (stefin A)	95
LU_S	1933_g_at	ABCC5	ATP-binding cassette, sub-family C (CFTR/MRP), member 5	95

Of the 12,533 probe sets interrogated by the oligonucleotide arrays, the 216 genes shown here were selected as predictors in the classification method.

Table 4. Predication scores in training and 'blinded' tumors

KEY:

= correct prediction = below confidence threshold = incorrect position

	PR	BL	BR	CO	GA	KI	LI	OV	PA	LU_A	LU_S	Dixon
PR29T	37.78	102.35	119.48	118.95	114.47	112.01	145.96	118.11	150.89	112.58	122.43	0.571
PR13BT	35.88	100.05	117.90	118.78	111.71	107.50	141.44	121.15	155.20	112.06	118.04	0.538
PR8T	35.73	100.49	109.00	113.73	113.12	111.53	138.86	114.74	147.39	104.37	117.41	0.580
PR21T	35.10	101.25	112.55	114.66	110.51	112.93	140.74	118.91	151.94	106.47	113.23	0.566
PR27T	40.44	106.62	111.35	122.05	115.52	119.98	146.58	123.98	153.24	104.90	116.92	0.571
PR9T	32.47	97.93	107.67	110.85	111.56	110.74	135.91	115.34	146.19	99.98	110.24	0.576
PR7T	36.09	97.30	103.78	113.22	110.32	107.49	135.73	115.13	145.57	99.56	110.93	0.559
PR19T	35.74	103.29	110.04	117.30	115.53	107.88	139.85	117.46	145.56	104.17	120.53	0.615
PR5T	32.88	96.09	112.97	118.80	111.05	112.64	141.88	118.52	151.26	106.41	115.95	0.534
PR24T	40.25	109.02	118.78	122.49	120.39	116.22	146.56	123.26	154.12	108.60	118.13	0.600
U40	44.15	107.49	115.20	119.75	113.86	121.02	144.44	120.80	159.30	106.04	120.28	0.537
U41	57.43	100.33	101.19	99.76	103.17	111.13	117.26	107.50	142.04	97.63	107.45	0.475
PR4	35.87	95.09	100.64	108.75	104.69	107.12	133.84	110.37	140.80	96.50	106.06	0.564
PR3	36.64	100.50	107.76	114.14	111.31	114.17	140.98	115.59	145.62	100.50	115.80	0.586
PR1	36.00	100.92	114.50	121.30	112.57	112.53	146.73	121.41	152.28	108.66	116.84	0.558
PR10	38.75	104.34	105.02	112.00	113.51	114.77	136.87	115.11	145.07	100.17	114.01	0.578
PR22	37.65	100.58	112.59	114.84	109.97	110.77	138.66	113.24	149.06	104.78	114.01	0.565
PR12	32.47	103.10	114.33	116.83	112.60	116.73	143.88	116.41	148.40	106.67	116.10	0.609
PR31	36.93	104.78	113.90	116.23	114.22	115.29	147.00	123.66	153.07	106.52	117.61	0.584
PR30	34.65	97.32	111.33	114.90	111.28	109.66	136.11	113.94	148.00	105.50	118.41	0.553
PR26	42.40	98.86	112.51	116.74	110.39	116.24	137.84	116.80	149.04	103.81	114.85	0.529
PR16	38.60	102.92	107.02	112.69	114.65	113.21	138.09	114.20	143.86	101.78	117.63	0.600
PR23	36.83	98.55	113.37	116.74	111.06	115.80	142.93	116.77	149.33	106.78	117.75	0.551
PR6	37.86	92.22	105.02	104.34	103.69	101.39	131.09	105.33	139.29	98.45	112.49	0.536
PR11	35.54	105.95	116.68	118.13	113.29	112.56	141.47	122.20	154.08	110.38	119.40	0.594
PR17	44.13	106.80	114.29	119.20	113.35	120.45	144.09	120.04	157.49	105.03	119.99	0.537
BL7T	97.17	53.84	92.00	94.77	88.99	91.16	120.49	95.30	127.85	87.24	88.78	0.451
BL1T	99.64	53.24	84.40	93.80	89.79	97.15	120.50	87.92	120.83	86.00	89.03	0.461
BL18T	107.92	70.38	79.99	86.32	78.99	91.73	115.73	86.80	113.65	81.62	73.50	0.069
BL19T	98.88	57.94	83.12	84.31	78.06	93.94	109.70	87.26	122.38	81.79	79.85	0.312
BL16T	109.68	65.89	86.34	87.84	88.35	93.64	111.45	91.66	130.53	87.95	87.01	0.316
BL10T	103.97	57.91	95.64	90.91	86.17	100.92	125.38	98.82	125.47	91.91	93.43	0.418
BL2T	97.29	58.20	75.82	88.13	85.70	77.68	107.54	81.69	112.78	73.90	87.45	0.288
BL9T	97.47	54.02	86.06	91.48	88.50	90.09	117.09	85.77	117.54	81.45	89.49	0.432
BR8T	117.88	88.42	57.94	81.94	84.83	91.67	104.04	84.60	119.11	80.25	87.39	0.365
BR10T	111.53	84.57	56.56	74.78	75.93	98.85	98.81	78.51	116.87	75.91	87.68	0.302
BR14T	113.10	91.41	62.02	86.00	94.06	99.13	106.99	86.78	120.84	81.05	97.80	0.324
BR16T	104.96	84.65	54.83	88.75	86.16	90.20	107.70	87.29	116.73	79.28	93.93	0.395
BR17T	109.97	90.70	57.52	97.34	91.07	97.94	121.15	97.21	123.48	85.04	98.93	0.417
BR20T	122.86	91.48	70.63	94.45	104.27	104.06	108.05	86.37	125.58	93.18	109.34	0.287
BR6T	109.35	78.57	56.46	85.43	85.39	86.19	105.30	80.97	110.16	71.95	83.80	0.288
BR15T	107.51	80.79	55.06	82.14	82.42	86.49	99.80	86.70	113.37	70.65	91.17	0.267
BR21T	112.33	79.14	57.52	83.06	85.54	90.07	101.12	78.78	117.05	73.78	75.60	0.273
BR29T	115.78	88.50	58.01	90.08	87.89	98.76	115.67	93.23	121.92	82.26	92.16	0.379
BR31T	108.34	81.29	59.76	84.46	82.73	93.76	104.94	82.16	116.60	74.59	79.22	0.261
BR32T	114.61	87.60	58.26	89.76	83.60	101.63	111.53	92.28	122.06	79.11	89.03	0.327

KEY:
= correct prediction = below confidence threshold = incorrect position

	PR	BL	BR	CO	GA	KI	LI	OV	PA	LU_A	LU_S	Dixon
U1	97.90	78.85	65.46	77.97	76.97	85.37	102.16	86.25	115.81	67.04	81.26	0.031
U16	118.60	94.30	84.13	91.78	103.83	105.06	105.13	86.22	125.70	84.80	94.35	0.016
UX7	110.80	85.08	77.69	89.95	83.04	89.29	106.00	88.76	124.12	80.66	87.72	0.064
UX8	103.92	85.51	65.91	90.02	85.91	79.06	110.13	92.49	118.08	77.74	88.88	0.227
UX19	101.86	80.46	61.30	87.86	87.45	86.70	106.87	85.35	115.36	77.43	93.34	0.298
B24T	106.93	83.37	56.09	85.53	85.40	97.99	104.18	88.27	118.69	79.20	86.04	0.369
B46T	112.27	84.58	75.13	79.40	78.00	94.61	104.63	90.03	123.84	82.68	83.56	0.059
B30T	108.56	83.52	58.86	87.62	81.69	90.25	107.76	92.17	120.36	75.45	84.11	0.270
B34T	109.22	87.41	59.00	85.47	81.89	93.29	107.69	91.31	118.66	79.58	91.84	0.345
B36T	116.02	88.88	64.98	103.47	89.42	95.66	122.74	97.79	129.50	87.08	99.06	0.343
B37T	103.07	78.95	59.68	81.59	76.21	84.81	103.06	83.73	120.27	75.32	76.45	0.258
B38T	108.41	81.27	57.44	96.93	89.83	87.75	116.02	92.83	121.69	79.13	97.09	0.338
B39T	107.41	76.23	57.59	78.07	75.74	87.25	96.27	74.32	107.62	67.55	76.53	0.199
B41T	103.41	79.07	57.69	80.04	72.15	91.54	105.55	83.22	113.25	71.30	83.30	0.245
CO7T	121.27	90.65	85.31	50.68	78.51	100.95	111.54	89.77	121.85	84.58	99.98	0.391
CO9T	113.91	85.48	83.78	57.33	76.90	98.24	117.61	86.70	109.63	79.63	94.55	0.325
CO42T	110.94	79.91	79.22	59.38	69.50	94.56	108.16	85.69	106.47	77.90	94.33	0.188
CO15T	124.61	99.85	93.56	59.41	92.14	111.41	111.89	94.44	126.76	97.92	107.59	0.486
CO40T	112.95	88.69	89.10	55.59	70.23	98.60	117.86	96.40	116.18	84.33	95.24	0.235
CO27T	116.62	92.25	84.90	54.10	77.85	98.05	100.87	88.52	115.83	85.31	96.75	0.380
CO30T	125.39	94.24	92.23	54.41	86.25	106.43	116.11	91.66	120.51	94.05	102.92	0.449
CO32T	111.45	86.30	91.06	57.24	68.47	99.12	120.88	92.67	115.93	89.68	100.35	0.177
CO20T	118.54	89.42	86.14	54.55	78.44	104.45	107.67	89.23	120.43	85.49	96.86	0.363
CO24T	113.70	89.70	82.71	57.18	80.51	94.54	105.67	87.21	113.09	82.05	98.02	0.413
CO8T	119.99	88.49	83.06	55.72	75.87	101.23	114.85	92.98	121.80	84.11	95.70	0.305
U6	115.11	90.07	90.58	62.64	71.08	97.31	120.02	100.83	117.81	85.46	102.14	0.147
U12	107.24	84.22	83.37	65.02	80.89	97.15	114.35	85.15	115.22	84.69	97.43	0.316
CO14T	116.05	91.14	93.22	54.84	75.29	102.42	116.04	96.50	118.80	91.42	98.70	0.320
CO21T	118.66	92.32	95.99	68.10	84.47	103.63	117.67	100.13	112.82	88.00	107.40	0.324
CO23T	116.23	85.92	98.68	64.72	79.10	101.22	119.90	96.90	122.47	94.47	101.02	0.249
CO5T	105.09	84.81	83.32	55.54	73.59	90.21	105.09	82.27	114.32	81.28	90.47	0.307
CO43T	108.78	86.92	90.50	65.09	78.61	91.43	113.82	93.96	116.48	86.47	93.61	0.263
CO44T	99.65	84.21	85.41	66.83	68.17	92.06	109.01	87.24	114.10	78.63	84.15	0.028
CO49T	126.02	105.60	112.95	71.80	86.12	113.38	135.61	116.39	121.33	107.76	114.35	0.224
CO51T	116.28	95.05	100.45	60.57	72.62	104.65	126.11	100.84	123.47	96.54	104.48	0.184
CO56T	116.25	92.82	100.79	61.04	71.51	106.30	127.14	100.19	125.18	96.19	101.11	0.158
CO61T	123.23	99.41	98.55	75.04	84.24	106.10	122.86	107.24	112.43	96.63	103.70	0.191
GA9T	109.87	81.42	95.15	87.63	66.20	100.15	119.39	99.93	127.31	93.00	96.01	0.249
GA46T	112.91	80.69	74.41	69.69	69.03	98.60	112.39	93.66	108.67	76.34	74.75	0.015
GA18T	112.96	84.17	86.72	81.27	63.32	103.00	111.27	93.48	129.54	90.76	97.73	0.271
GA8T	113.75	91.13	98.57	82.22	57.36	106.49	126.63	107.00	130.21	93.10	97.87	0.341
GA3T	118.15	86.51	85.24	76.04	62.73	99.54	117.29	100.10	117.42	89.22	100.61	0.240
GA2T	101.31	74.03	86.74	71.80	55.58	91.69	115.68	96.24	116.57	81.59	91.23	0.266
GA5T	112.13	89.47	95.36	75.75	58.81	102.83	123.40	98.19	126.29	93.15	103.55	0.251
GA6T	112.66	91.23	88.55	97.89	72.84	100.65	121.82	103.74	133.90	84.16	92.44	0.185
GA116X	122.22	86.08	81.08	67.08	64.32	95.38	105.81	89.53	115.91	84.57	102.78	0.048
GA280	105.18	79.81	71.78	70.90	67.85	89.35	101.93	87.22	107.36	65.73	77.61	0.051
GA102X	127.08	96.63	94.37	77.52	66.66	112.02	121.20	95.58	129.95	96.38	107.39	0.171

KEY:
= correct prediction = below confidence threshold = incorrect position

	PR	BL	BR	CO	GA	KI	LI	OV	PA	LU_A	LU_S	Dixon
U3	110.46	84.63	72.55	78.75	88.59	92.67	97.04	79.15	114.32	70.90	76.68	0.038
UX10	104.73	74.44	81.29	80.84	73.81	85.36	110.53	84.17	111.28	78.34	86.86	0.017
KI1T	113.49	95.04	99.12	108.35	101.78	41.73	125.25	107.50	131.36	97.47	114.88	0.595
KI20T	115.68	91.81	89.36	95.88	100.64	46.05	109.24	95.53	125.87	92.24	105.06	0.543
KI18T	109.59	89.35	95.88	103.88	98.83	41.84	121.12	105.75	127.93	90.70	105.09	0.552
KI3T	112.47	99.51	107.79	110.58	105.07	49.02	132.05	111.60	135.17	101.64	114.00	0.586
KI22T	107.29	90.37	101.73	104.97	98.87	41.80	121.98	104.40	128.15	95.66	110.37	0.562
KI2T	110.80	87.17	91.07	99.94	96.18	42.97	116.28	99.88	122.95	87.35	102.47	0.553
KI4T	110.37	92.76	88.09	93.57	101.79	46.76	110.15	97.48	123.87	89.77	107.10	0.536
KI19T	110.73	90.44	85.61	95.08	98.51	46.40	106.93	99.09	125.03	89.11	103.31	0.499
KI17T	109.61	87.66	94.53	95.26	92.16	59.67	104.28	100.07	122.78	87.26	102.86	0.437
KI16T	116.52	95.63	95.81	100.01	104.60	46.75	118.35	100.24	127.34	100.20	110.53	0.607
UX14	102.42	81.32	82.02	88.70	77.98	75.43	107.98	89.86	116.93	77.65	88.00	0.054
LI34T	118.66	90.23	72.14	80.08	90.02	93.25	77.48	86.69	117.35	82.43	96.43	0.115
LI11T	150.88	122.37	110.80	125.20	121.15	119.15	56.33	127.71	160.06	116.07	129.36	0.525
LI32T	146.78	123.36	115.76	121.88	122.21	124.34	55.24	123.56	159.89	118.11	135.21	0.578
LI30T	140.70	116.44	106.49	109.22	114.26	118.38	62.87	115.85	150.48	108.91	118.60	0.498
LI35T	148.50	116.54	114.18	110.44	120.88	129.64	64.40	111.77	156.73	120.34	124.62	0.499
LI13T	143.88	127.11	124.06	127.63	127.87	115.09	64.59	132.89	162.79	120.87	131.83	0.514
U9	112.53	98.06	90.17	91.95	101.13	100.25	100.71	90.15	135.87	88.18	106.75	0.041
OV23T	118.26	92.41	82.22	90.65	96.77	98.49	115.63	58.10	122.74	85.78	98.57	0.373
OV1AT	123.70	89.18	89.45	99.20	99.83	109.58	131.77	65.12	123.21	94.24	104.29	0.361
OV16T	117.12	99.03	92.14	94.06	101.35	99.65	116.65	61.37	133.46	102.14	114.14	0.427
OV27T	127.78	89.58	85.97	91.58	98.81	106.65	114.67	63.59	128.92	95.11	101.25	0.343
OV3T	115.95	89.11	91.08	80.12	89.88	103.34	114.42	64.38	125.17	94.35	91.71	0.259
OV2AT	118.83	81.61	80.13	97.22	96.11	99.75	121.75	60.15	118.23	90.20	95.18	0.324
OV21T	128.23	96.64	88.52	90.34	107.91	110.46	106.95	62.22	128.16	101.60	104.76	0.398
OV8T	104.65	77.14	76.12	78.21	81.46	87.17	100.94	69.81	114.33	82.17	92.51	0.142
OV7T	116.04	86.31	90.18	92.45	96.69	101.66	122.64	58.68	123.74	93.12	101.64	0.425
U7	112.44	88.28	83.82	98.00	100.92	100.10	127.48	65.47	121.75	89.63	100.04	0.296
U8	119.08	90.88	89.94	98.37	92.97	100.81	122.52	66.56	127.42	87.57	96.69	0.345
U11	106.24	77.83	90.00	96.24	94.19	96.25	124.78	66.34	124.62	86.74	97.28	0.197
UX20	115.87	89.18	83.85	89.80	97.04	103.13	121.62	62.65	124.20	91.20	93.84	0.344
OVR16	107.71	78.16	82.45	91.46	89.73	91.62	120.97	64.55	118.98	84.26	94.26	0.241
OVR19	109.02	82.81	93.99	93.05	92.92	94.60	124.23	57.69	124.95	93.04	103.88	0.373
OVR27	106.96	78.11	78.21	90.74	87.96	93.66	119.22	63.86	116.66	81.26	87.11	0.257
OVR28	110.48	87.31	90.78	92.09	96.77	102.13	127.17	72.09	135.15	88.23	102.43	0.241
OVR2	112.35	79.81	83.93	96.13	91.18	100.07	127.39	64.36	122.19	87.21	96.63	0.245
OVR5	104.12	72.61	78.18	82.00	85.44	95.44	110.24	74.76	120.52	81.70	84.22	0.045
OVR8	109.15	88.15	90.55	91.04	91.92	101.71	124.23	65.85	132.16	92.50	95.88	0.336
OVR11	118.28	88.61	89.35	93.88	96.58	100.83	124.69	59.25	127.53	91.29	97.23	0.430
OVR12	110.75	82.40	79.36	92.78	91.36	97.10	122.04	68.42	114.36	86.12	94.77	0.204
OVR13	123.95	95.67	102.85	108.56	103.16	110.48	142.22	73.24	142.82	104.71	109.38	0.322
OVR22	113.92	84.46	79.98	94.58	96.31	103.54	119.95	60.42	117.11	84.46	93.32	0.329
OVR26	115.44	90.12	84.45	97.96	103.96	103.50	120.81	61.92	129.09	94.33	106.36	0.335
OVR10	108.95	84.66	80.67	83.32	90.42	90.32	107.24	61.71	126.21	88.72	99.32	0.294
OVR1	115.86	93.33	97.10	102.08	100.03	101.25	130.69	66.44	133.76	95.43	108.97	0.399
PA16BT	158.23	126.98	125.58	125.12	125.90	137.24	161.51	128.88	62.06	121.83	132.43	0.601

KEY:
= correct prediction = below confidence threshold = incorrect position

	PR	BL	BR	CO	GA	KI	LI	OV	PA	LU_A	LU_S	Dixon
PA8T	147.25	122.30	118.38	113.14	120.42	124.14	149.87	127.65	59.28	116.53	135.01	0.595
PA11T	151.78	126.70	122.66	119.92	117.88	127.97	154.77	126.37	59.87	118.29	131.23	0.611
PA22T	153.33	132.98	128.12	124.12	130.28	134.31	155.20	130.97	68.14	131.26	145.05	0.643
PA23T	129.37	94.37	91.97	89.13	98.19	101.86	126.58	96.64	79.50	89.22	106.78	0.193
PA17T	166.25	125.64	125.67	132.54	138.37	137.33	161.31	137.30	80.01	126.49	129.16	0.529
LU_A31T	108.44	91.50	85.10	92.88	98.35	97.04	115.87	98.47	127.48	60.33	88.91	0.369
LU_A34T	108.02	84.64	76.70	83.89	84.08	94.86	112.14	90.50	114.63	54.49	80.56	0.369
LU_A39T	88.64	75.06	77.82	79.65	70.82	84.83	105.37	90.73	117.60	62.83	71.73	0.146
LU_A8T	110.82	86.66	83.38	91.51	88.83	98.64	117.78	101.88	120.85	62.65	88.98	0.356
LU_A20T	110.22	89.77	83.41	96.76	92.79	103.19	118.71	107.77	127.70	64.51	90.50	0.299
LU_A5T	109.51	79.20	81.07	85.84	79.17	89.68	104.40	90.20	115.73	68.08	74.02	0.125
LU_A6T	114.71	88.98	69.47	83.13	94.12	95.32	102.91	85.93	113.90	64.64	90.04	0.096
LU_A17T	100.71	75.84	67.28	76.38	79.39	83.55	106.02	80.92	102.29	55.84	76.33	0.228
LU_A18T	104.68	86.57	85.44	84.70	86.05	90.93	113.67	91.47	112.35	61.20	87.86	0.448
U17	105.70	70.20	81.43	89.40	82.67	90.74	114.34	89.68	127.72	71.32	87.81	0.020
UX4	107.98	87.74	76.15	78.97	87.94	92.87	109.52	88.63	115.56	65.13	79.16	0.219
LU40T	107.23	87.64	81.67	93.19	95.01	93.99	119.01	97.82	118.00	62.93	91.42	0.334
LU44T	114.33	89.09	88.75	95.24	94.28	91.81	113.31	95.14	118.38	65.57	76.19	0.201
LU33T	113.24	88.48	87.12	95.41	97.56	97.39	117.94	95.56	121.10	65.47	90.89	0.389
LU_S19T	98.95	79.88	81.32	85.73	83.75	94.58	114.63	89.43	120.15	69.30	67.65	0.031
LU_S24T	120.23	87.84	98.82	105.57	95.98	111.93	126.19	107.57	134.64	89.65	59.00	0.381
LU_S25T	110.21	86.09	94.86	97.68	95.89	110.77	128.35	102.05	128.24	80.09	59.75	0.297
LU_S12T	117.58	87.70	91.33	105.01	97.52	111.67	129.97	105.96	131.69	91.00	60.11	0.385
LU_S13T	112.44	81.45	68.24	78.34	87.64	94.33	94.54	76.06	120.32	75.45	84.83	0.138
LU_S14T	120.80	87.78	97.02	104.17	104.87	113.00	129.14	103.10	130.68	80.65	64.41	0.245
LU_S11T	127.67	90.70	98.78	106.95	100.16	114.54	131.14	112.48	137.39	89.48	60.05	0.381
LU_S7T	122.35	86.60	94.17	104.52	91.89	110.79	125.14	108.48	134.99	90.62	65.12	0.307
U2	118.22	90.71	97.79	100.68	93.78	114.54	129.42	110.06	136.91	85.28	57.73	0.348
U19	128.13	92.29	98.84	109.28	103.27	113.43	128.98	111.45	132.97	90.33	64.08	0.381
LU41T	121.67	89.04	95.55	101.55	95.02	114.58	126.67	105.54	135.31	87.16	58.76	0.371
LU30T	121.91	94.49	101.04	103.95	99.32	118.57	130.95	108.88	142.40	95.64	60.72	0.413
LU26T	125.21	89.51	94.09	94.53	98.91	109.23	119.91	100.96	113.42	90.33	74.84	0.291
LU36T	116.60	76.85	83.54	95.23	93.30	97.76	116.67	90.89	117.01	75.12	67.04	0.162

Shown is the sum of absolute difference scores for each prediction in each tumor. The Dixon confidence score is shown in the right-hand column. A score >0.1 indicates a confident prediction.

[0136] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference for all purposes.

We Claim:

1. A kit for identifying an origin of a tumor in a subject, wherein the tumor is of prostate, breast, colorectal, lung, ovarian, gastroesophageal, pancreatic, liver, kidney or bladder origin, the kit comprising:

a) a probe that can detect an expression product of a gene in a first tumor class as indicated in Table 3; and

b) a probe that can detect an expression product of a gene in a second tumor class as indicated in Table 3.

2. The kit of claim 1, wherein the kit comprises at least two probes that can detect an expression product of genes in the first tumor class.

3. The kit of claim 1, wherein the kit comprises at least two probes that can detect an expression product of genes in the second tumor class.

4. The kit of claim 1, wherein the kit comprises at least two probes that can detect an expression product of genes in the first tumor class and at least two probes that can detect an expression product of genes in the second tumor class.

5. The kit of claim 1, wherein the kit further comprises: c) at least a third probe that can detect an expression product of a gene in at least a third tumor class as indicated in Table 3.

6. The kit of claim 5, wherein the kit comprises ten probes, each of which can detect an expression product of a gene in a different tumor class as indicated in Table 3.

7. The kit of claim 1, wherein the expression product is an mRNA transcribed from the gene.

8. The kit of claim 7, wherein the probes are oligonucleotides that can hybridize to the mRNA, or to a cDNA or cRNA copy of the mRNA.

9. The kit of claim 8, wherein the oligonucleotides are attached to a solid support.

10. The kit of claim 9, wherein the solid support comprises a microchip.

11. The kit of claim 1, wherein the expression product is a polypeptide encoded by the gene.

12. The kit of claim 11, wherein the probes each comprise an antibody.

13. The kit of claim 12, wherein the antibodies are monoclonal antibodies.

14. The kit of claim 11, wherein the probes are attached to a solid support.

15. A method for identifying an origin of a tumor, the method comprising detecting in a tumor sample an expression level of at least two genes, each of which genes is diagnostic for a different tumor class as identified in Table 3, wherein an elevated level expression for a gene indicates that the tumor originated from the tumor class for which the gene is diagnostic.

16. The method or claim 15, wherein the tumor is of prostate cancer, breast cancer, colorectal cancer, lung adenocarcinoma, lung squamous cell carcinoma, ovarian cancer, gastroesophageal cancer, pancreatic cancer, liver cancer, kidney cancer or bladder cancer origin.

17. The method of claim 15, wherein an expression level is determined for at least three genes, each of which genes is diagnostic for a different tumor class as identified in Table 3.

18. The method of claim 15, wherein an expression level is determined for at least two genes that are both diagnostic for a single tumor class as identified in Table 3.

19. The method of claim 15, wherein an expression level is determined for at least two genes that are both diagnostic for a first tumor class as identified in Table 3, and at least two genes that are both diagnostic for a second tumor class as identified in Table 3.

20. The method of claim 15, wherein an expression level is determined for at least three genes in each of two or more tumor classes as identified in Table 3.

21. The method of claim 15, wherein an expression level is determined for one or more genes in each of at least ten tumor classes identified in Table 3.

22. The method of claim 15, wherein the expression level is elevated compared to expression level of the gene in a non-cancer control sample.

23. The method of claim 15, wherein the expression level is elevated compared to expression of the gene in a control sample obtained from a tumor of a different tumor class.

24. The method of claim 15, wherein the expression level of a gene is determined by detecting the level of expression of an mRNA transcribed from the gene.

25. The method of claim 24, wherein the level of expression of mRNA is detected by techniques selected from the group consisting of northern blot analysis, reverse transcriptase PCR, real time quantitative PCR and hybridization to an oligonucleotide array.

26. The method of claim 15, wherein the expression level of a gene is determined by detecting the level of expression of a protein encoded by the gene.

27. The method of claim 26, wherein the level of expression of the protein is detected through western blotting or an array by utilizing a labeled probe specific for the protein.

28. The method of claim 27, wherein the probe is an antibody.

29. The method of claim 28, wherein the antibody is a monoclonal antibody.

30. The method of claim 15, wherein the tumor is a metastatic lesion or a primary tumor.

31. A method for identifying an origin of a tumor, the method comprising:

a) providing a predictor set that comprises expression levels for two or more genes, each of which is diagnostic for a different tumor class as identified in Table 3;

b) detecting in a tumor sample an expression level of at one gene that is diagnostic for a tumor class as identified in Table 3; and

c) calculating a vector distance from the expression level obtained from the tumor sample to each of the expression levels of the predictor set,

wherein the shortest vector distance indicates the origin of the tumor.

32. The method of claim 31, wherein the predictor set comprises expression levels for at least three genes, each of which genes is diagnostic for a different tumor class as identified in Table 3.

33. The method of claim 31, wherein the predictor set comprises expression levels for at least two genes that are both diagnostic for a single tumor class as identified in Table 3.

34. The method of claim 31, wherein the predictor set comprises expression levels for at least two genes that are both diagnostic for a first tumor class as identified in Table 3, and at least two genes that are both diagnostic for a second tumor class as identified in Table 3.

35. The method of claim 31, wherein the predictor set comprises expression levels for at least three genes in each of two or more tumor classes as identified in Table 3.

36. The method of claim 31, wherein the predictor set comprises expression levels for one or more genes in each of at least ten tumor classes identified in Table 3.

37. The method of claim 31, wherein the expression level of a gene in the tumor sample is determined by detecting the level of expression of an mRNA transcribed from the gene.

38. The method of claim 31, wherein the expression level of a gene in the tumor sample is determined by detecting the level of expression of a protein encoded by the gene.

39. The method of claim 31, wherein the tumor sample is obtained from a metastatic lesion or a primary tumor.

40. The method of claim 31, wherein the Dixon threshold for the shortest vector distance is 0.5 or less.

41. The method of claim 40, wherein the Dixon threshold for the shortest vector distance is 0.1 or less.

42. A method for obtaining a predictor set for classifying a sample into one of two or more classes, the method comprising:

a) obtaining a value for one or more features for each of a plurality of members of each of the classes;

b) determining a Wilcoxon rank score for each of the features to eliminate nonpredictive features; and

c) ranking the remaining features by predictive accuracy using a support vector machine.

43. The method of claim 42, wherein the features are genes and the values are expression levels of the genes.

44. The method of claim 42, wherein the classes are tumor classes.

45. The method of claim 42, wherein the classes are exposure of a sample to different conditions.

46. The method of claim 45, wherein the different conditions are exposure to different chemical compounds.

47. The method of claim 42, wherein the classes are different disease states.

48. The method of claim 42, wherein the method further comprises classifying a sample into one of the classes by:

a) determining a value for one or more features in the sample; and

b) calculating a vector distance from the obtained for the feature in the sample to each of the expression levels of the predictor set,

wherein the shortest vector distance indicates the class of which the sample is a member.

49. A method for screening a subject for prostate cancer or at risk of developing prostate cancer, the method comprising:

a) detecting a level of expression of at least one gene in a sample of prostate tissue obtained from the subject to provide a first value, wherein the gene is selected from the group consisting of LIM, multidrug resistance-associated protein homolog (MRP4), T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I; and

b) comparing the first value with a level of expression of the gene in a sample of prostate tissue obtained from a disease-free subject, wherein a greater expression level in the subject sample compared to the sample from the disease-free subject is indicative of the subject having prostate cancer or at risk of developing prostate cancer.

50. The method of claim 49, wherein the level of expression of at least two genes is detected.

51. The method of claim 49, wherein the level of expression of the gene is determined by detecting the level of expression of an mRNA corresponding to the gene.

52. The method of claim 51, wherein the level of expression of mRNA is detected by techniques selected from the group consisting of northern blot analysis, reverse transcriptase PCR, real time quantitative PCR and oligonucleotide arrays.

53. The method of claim 49, wherein the level of expression of the gene is determined by detecting the level of expression of a protein encoded by the gene.

54. The method of claim 53, wherein the level of expression of the protein is detected through western blotting or an array by utilizing a labeled probe specific for the protein.

55. The method of claim 54, wherein the probe is an antibody.

56. The method of claim 55 wherein the antibody is a monoclonal antibody.

57. A method for screening a subject for ovarian cancer or at risk of developing ovarian cancer, the method comprising:

a) detecting a level of expression of at least one gene in a sample of ovarian tissue obtained from the subject to provide a first value, wherein the gene is selected from the group consisting of laminin, alpha 5; vacuolar proton pump, beta polypeptide; putative cytoskeletal protein, natriuretic peptide receptor A, eyes absent homolog, U90916, AL049313, S100 alpha, keratinocyte transglutaminase, GPCR64, meis1, spondin 1, GPCR39, AL050069, mammoglobin 2, and branched chain aminotransferase 1, cytosolic; mesothelin and kallikrein 6.

b) comparing the first value with a level of expression of the gene in a sample of ovarian tissue obtained from a disease-free subject, wherein a greater expression level in the subject sample compared to the sample from the disease-free subject is indicative of the subject having ovarian cancer or at risk of developing ovarian cancer.

58. The method of claim 57, wherein the level of expression of at least two genes is detected.

59. The method of claim 57, wherein the level of expression of the gene is determined by detecting the level of expression of a mRNA corresponding to the gene.

60. The method of claim 57, wherein the level of expression of mRNA is detected by techniques selected from the group consisting of northern blot analysis, reverse transcriptase PCR, real time quantitative PCR and oligonucleotide arrays.

61. The method of claim 57, wherein the level of expression of the gene is determined by detecting the level of expression of a protein encoded by the gene.

62. The method of claim 61, wherein the level of expression of the protein is detected through western blotting or an array by utilizing a labeled probe specific for the protein.

63. The method of claim 61, wherein the probe is an antibody.

64. The method of claim 63, wherein the antibody is a monoclonal antibody.

65. A method for monitoring the progression of prostate cancer in a subject having, or at risk of having a prostate cancer, the method comprising:

a) measuring a level of expression of at least one gene selected from the group consisting of LIM, multidrug resistance-associated protein homolog (MRP4), T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I, in a prostate tissue sample obtained from the subject, wherein an increase in the level of expression of the gene over time is indicative of the progression of the prostate cancer in the tissue.

66. A method for monitoring the progression of ovarian cancer in a subject having, or at risk of having, an ovarian cancer, the method comprising:

a) measuring a level of expression of at least one gene selected from the group consisting of laminin, alpha 5; vacuolar proton pump, beta polypeptide; putative cytoskeletal protein, natriuretic peptide receptor A, eyes absent homolog, U90916, AL049313, S100 alpha, keratinocyte transglutaminase, GPCR64, meis1, spondin 1, GPCR39, AL050069, mammoglobin 2, and branched chain aminotransferase 1, cytosolic, in

an ovarian tissue sample obtained from the subject, wherein an increase in the level of expression of the gene over time is indicative of the progression of the ovarian cancer in the tissue.

67. A method for identifying agents for use in treatment of prostate cancer comprising:

a) contacting a sample of diseased prostate cells with a candidate agent;

b) detecting a level of expression of at least one gene in the diseased prostate cells, wherein the gene is selected from the group consisting of LIM, multidrug resistance-associated protein homolog (MRP4), T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I; and

c) comparing the level of expression of the gene in the sample in the presence of the candidate agent with a level of expression of the gene in cells that are not contacted with the candidate agent, wherein a decreased level of expression of the gene in the sample in the presence of the candidate agent relative to the expression of the gene in the sample in the absence of the candidate agent is indicative of an agent useful in the treatment of prostate cancer.

68. A method for identifying agents for use in treatment of ovarian cancer comprising:

a) contacting a sample of diseased ovarian cells with a candidate agent;

b) detecting a level of expression of at least one gene in the diseased ovarian cells, wherein the gene is selected from the group consisting of laminin, alpha 5; vacuolar proton pump, beta polypeptide; putative cytoskeletal protein, natriuretic peptide receptor A, eyes absent homolog, U90916, AL049313, S100 alpha, keratinocyte transglutaminase, GPCR64, meis1, spondin 1, GPCR39, AL050069, mammoglobin 2, and branched chain aminotransferase 1, cytosolic; and

c) comparing the level of expression of the gene in the sample in the presence of the candidate agent with a level of expression of the gene in cells that are not contacted with the candidate agent, wherein a decreased level of expression of the gene in the sample in the presence of the candidate agent relative to the expression of the gene in the

sample in the absence of the candidate agent is indicative of an agent useful in the treatment of ovarian cancer.

69. A method of inhibiting undesired proliferation of a prostate cell, the method comprising administering to the cell an effective amount of an agent that can decrease the expression of at least one gene selected from the group consisting of of LIM, multidrug resistance-associated protein homolog (MRP4), T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I.

70. The method of claim 69, wherein the agent is selected from the group consisting of antisense nucleotides, ribozymes and double stranded RNAs.

71. A method of inhibiting undesired proliferation of an ovarian cell, the method comprising administering to the cell an effective amount of an agent that can decrease the expression of at least one gene selected from the group consisting of laminin, alpha 5; vacuolar proton pump, beta polypeptide; putative cytoskeletal protein, natriuretic peptide receptor A, eyes absent homolog, U90916, AL049313, S100 alpha, keratinocyte transglutaminase, GPCR64, meis1, spondin 1, GPCR39, AL050069, mammoglobin 2, and branched chain aminotransferase 1, cytosolic.

72. The method of claim 71, wherein the agent is selected from the group consisting of antisense nucleotides, ribozymes and double stranded RNAs.

73. A method for monitoring the efficacy of a treatment of a subject having prostate cancer or at risk of developing prostate cancer with an agent, the method comprising:

a) obtaining a pre-administration sample from the subject prior to administration of the agent:

b) detecting a level of expression of at least one gene selected from the group consisting of

c) LIM, multidrug resistance-associated protein homolog (MRP4), T-cell receptor Ti rearranged gamma-chain, testican, AC005053 and cam kinase I, in a preadministration sample;

d) obtaining one or more post-administration samples from the subject:

e) detecting a level of expression of the least one gene in the post-administration sample or samples;

f) comparing the level of expression of the gene in the pre-administration sample with the level of expression of the gene in the post-administration sample; and

g) adjusting the administration of the agent accordingly.

74. A method for monitoring the efficacy of a treatment of a subject having ovarian cancer or at risk of developing ovarian cancer with an agent, the method comprising:

a) obtaining a pre-administration sample from the subject prior to administration of the agent:

b) detecting a level of expression of at least one gene selected from the group consisting of

c) of laminin, alpha 5; vacuolar proton pump, beta polypeptide; putative cytoskeletal protein, natriuretic peptide receptor A, eyes absent homolog, U90916, AL049313, S100 alpha, keratinocyte transglutaminase, GPCR64, meis1, spondin 1, GPCR39, AL050069, mammoglobin 2, and branched chain aminotransferase 1, cytosolic, in the pre-administration sample;

d) obtaining one or more post-administration samples from the subject:

e) detecting a level of expression of the least one gene in the post-administration sample or samples;

f) comparing the level of expression of the gene in the pre-administration sample with the level of expression of the gene in the post-administration sample; and

g) adjusting the administration of the agent accordingly.

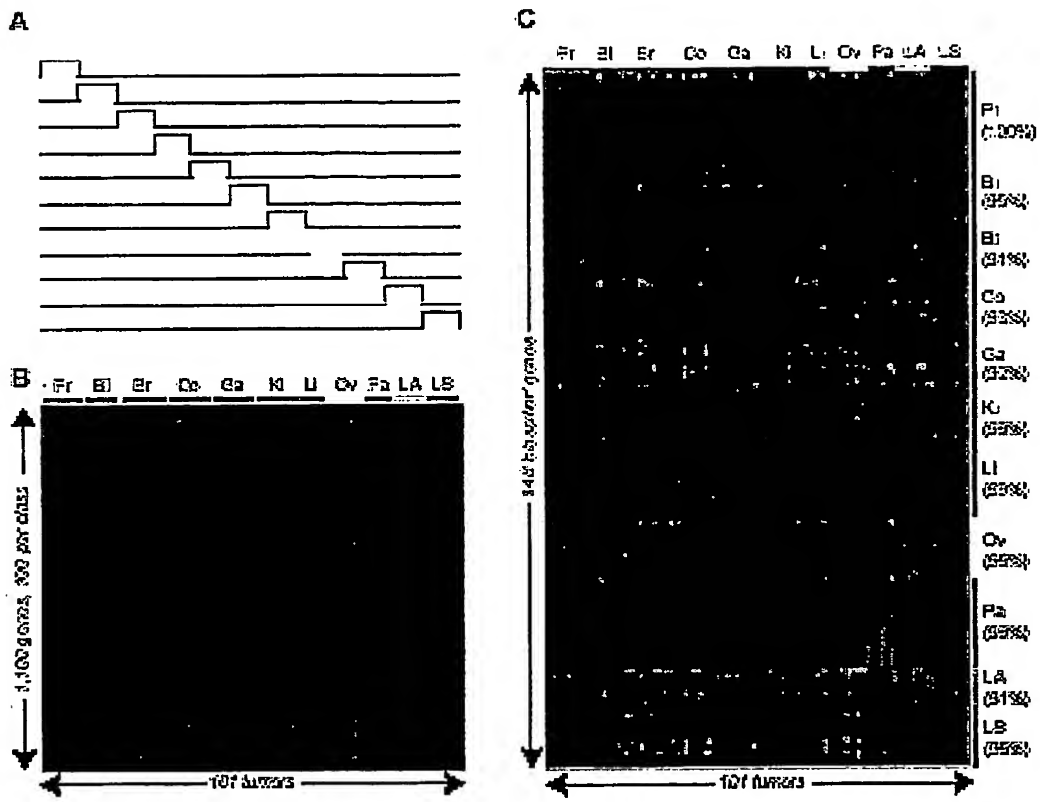
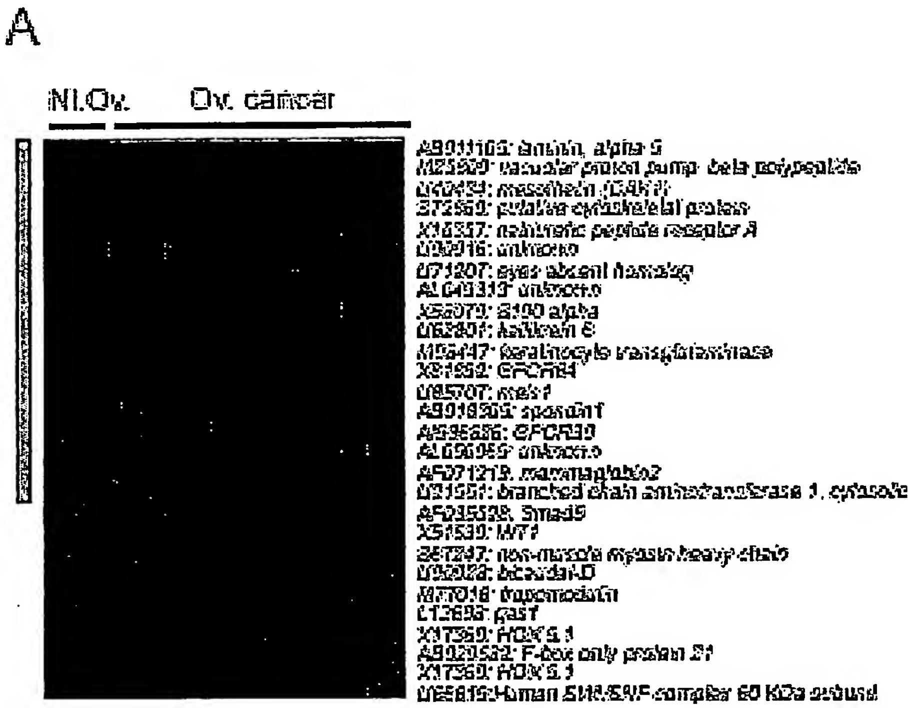


Figure 1


A
Nl.Ov.
Ov. cancer
AB011105: laminin, alpha 5
M25809: vacuolar proton pump beta polypeptide
U40434: mesothelin (CAK1)
Z72569: putative cytoskeletal protein
X16357: natriuretic peptide receptor A
AL049313: unknown
X58079: S100 alpha
U62801: kallikrein 6
M55447: keratinocyte transglutaminase
U85707: mei-1
AL050965: unknown
AF027219: mammaglobin2
U21551: branched chain aminotransferase 1, cytosolic
X51630: WT1
U12695: gas1
X17360: HOX 5.1
AB020522: F-box only protein 21
X17360: HOX 5.1
U66615: Human SWI/SNF complex 60 KDa subunit

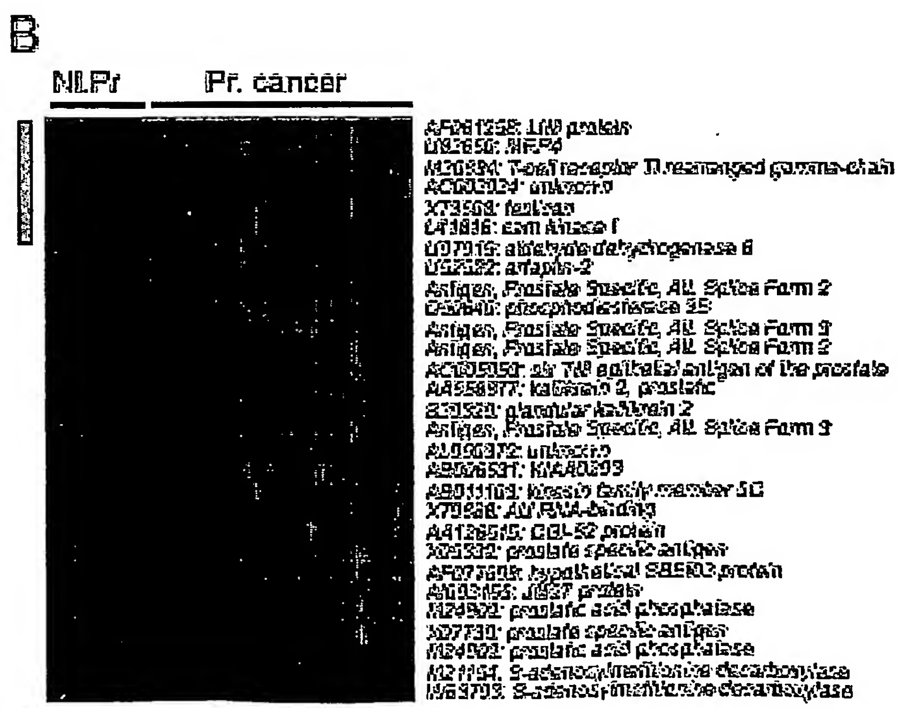

B
Nl.Pr
Pr. cancer
AC003024: unknown
U07919: aldehyde dehydrogenase 6
Antigen, Prostate Specific, Alt. Splice Form 2
Antigen, Prostate Specific, Alt. Splice Form 3
Antigen, Prostate Specific, Alt. Splice Form 2
AA558877: kallikrein 2, prostatic
Antigen, Prostate Specific, Alt. Splice Form 3
AL050372: unknown
X05332: prostate specific antigen
M24902: prostatic acid phosphatase
X07730: prostate specific antigen
M24902: prostatic acid phosphatase
M21154: S-adenosylmethionine decarboxylase

Figure 2

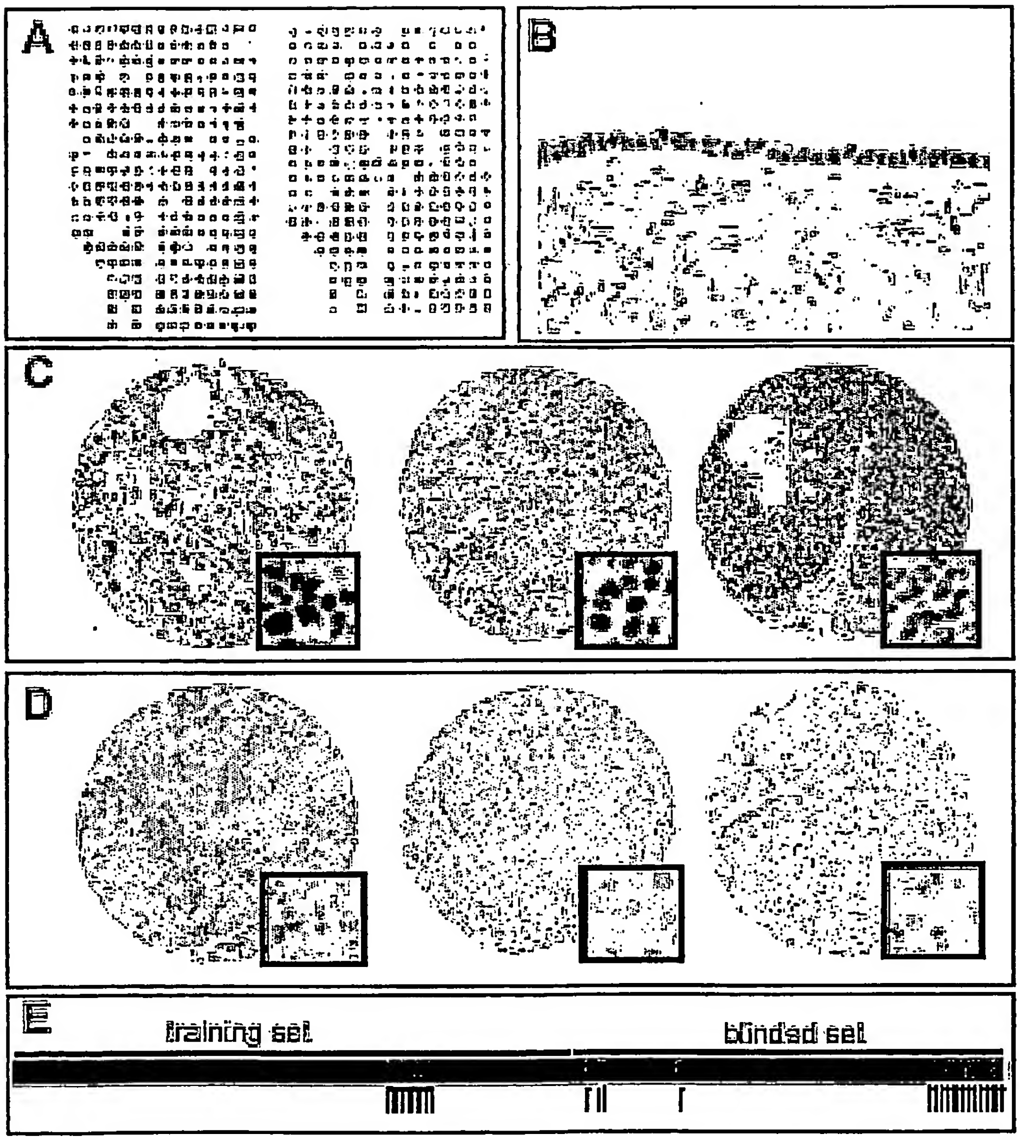


Figure 3

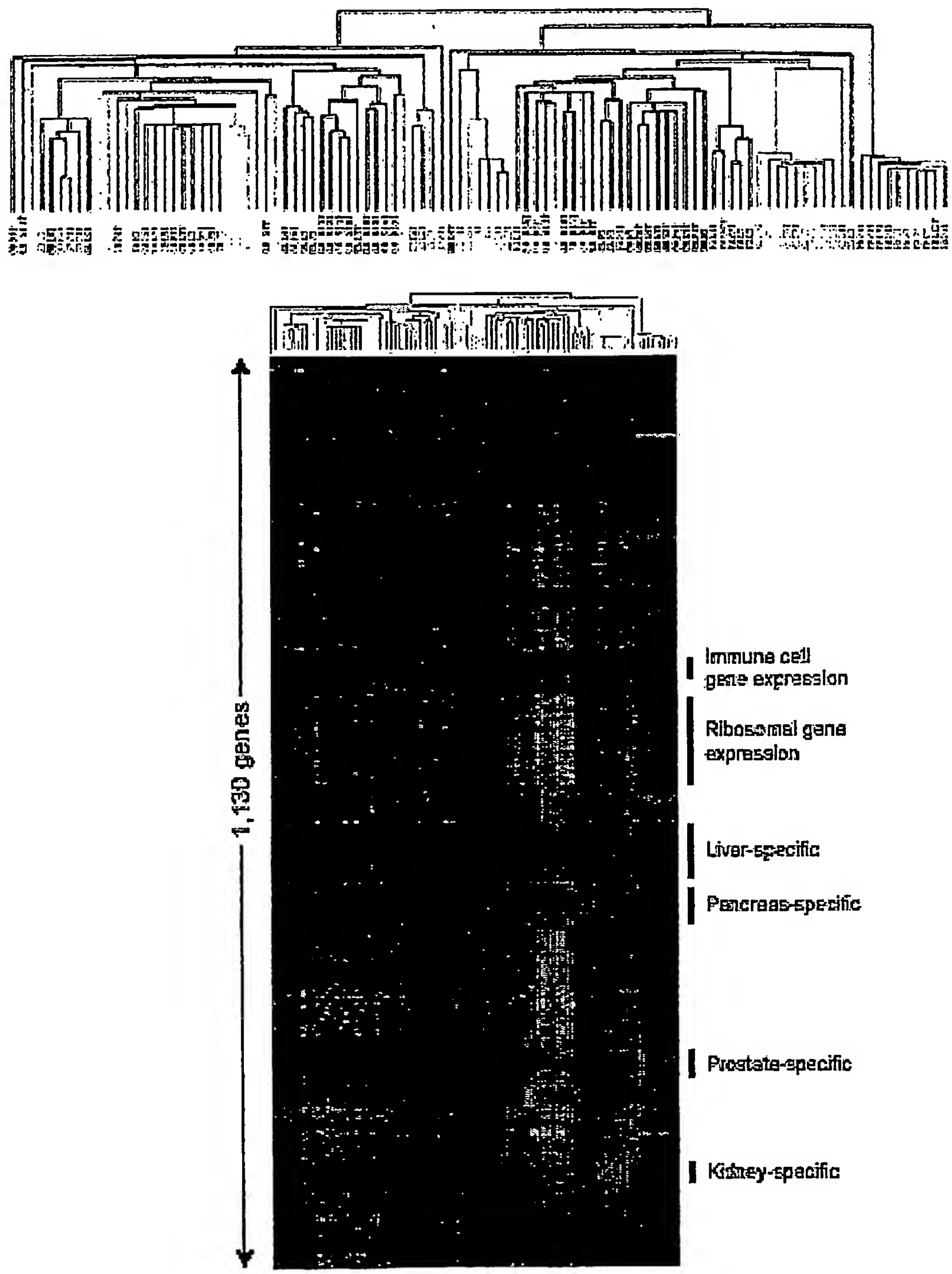

1,130 genes
Immune cell gene expression
Ribosomal gene expression
Liver-specific
Pancreas-specific
Prostate-specific
Kidney-specific

Figure 4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 April 2005 (14.04.2005)

PCT

(10) International Publication Number
WO 2005/032495 A2

(51) International Patent Classification[7]: A61K

(21) International Application Number: PCT/US2004/034163

(22) International Filing Date: 1 October 2004 (01.10.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/508,355 3 October 2003 (03.10.2003) US

(71) Applicant *(for all designated States except US)*: **BAYER PHARMACEUTICALS CORPORATION** [US/US]; 400 Morgan Lane, West Haven, Connecticut 06516 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TAYLOR, Ian** [US/US]; 149 Country Way, Madison, Connecticut 06443 (US). **PAULOSKI, Nicole, R.** [US/US]; 14 Aceto Street, Branford, Connecticut 06405 (US). **BIGWOOD, Douglas** [US/US]; 72 Kelsey Springs Drive, Madison, Connecticut 06443 (US).

(74) Agents: **GREENMAN, Jeffrey, M.** et al.; Bayer Pharmaceutuicals Corporation, 400 Morgan Lane, West Haven, Connecticut 06516 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(48) Date of publication of this corrected version: 16 June 2005

(15) Information about Correction:
see PCT Gazette No. 24/2005 of 16 June 2005, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: GENE EXPRESSION PROFILES AND METHODS OF USE

(57) Abstract: The present invention relates to gene expression profiles for lung cancer, microarrays comprising nucleic acid sequences representing gene expression profiles, and methods of using expression profiles and microarrays. The invention also provides methods and compositions for diagnostic assays for detecting cancer and therapeutic methods and compositions for treating cancer. The invention also provides methods for designing, identifying, and optimizing therapeutics for cancer.

WO 2005/032495 A2

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 December 2002 (19.12.2002)

PCT

(10) International Publication Number
WO 2002/101357 A3

(51) International Patent Classification[7]: C12Q 1/68, C07H 21/02, 21/04

(21) International Application Number: PCT/US2002/018628

(22) International Filing Date: 10 June 2002 (10.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/297,277 10 June 2001 (10.06.2001) US

(71) Applicant *(for all designated States except US)*: IRM LLC [US/—]; c/o Sofia House, 48 Church Street, HM 12 Hamilton (BM).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: SU, Andrew, I. [US/US]; 8236 Caminito Modena, La Jolla, CA 92037 (US). HAMPTON, Garret, M. [IE/US]; 5659 Arbor Grove Court, San Diego, CA 92121 (US).

(74) Agent: SMITH, Timothy, L.; Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, Suite E225, San Diego, CA 92121 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report: 5 August 2004

(15) Information about Correction:
Previous Correction:
see PCT Gazette No. 07/2004 of 12 February 2004, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2002/101357 A3

(54) Title: MOLECULAR SIGNATURES OF COMMONLY FATAL CARCINOMAS

(57) Abstract: This invention provides methods, kits, and algorithms for obtaining molecular signatures of cells based on their gene expression profiles. Devices for carrying out molecular signature analysis of unknown samples are also provided.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/18628

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C07H 21/02, 21/04

US CL : 435/6; 536/23.1, 24.3

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 536/23.1, 24.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WEST, PubMed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A; E	US 6,500,938 B1 (AU-YOUNG et al.), 31 December 2002, whole document	1-30

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
14 May 2003 (14.05.2003)

Date of mailing of the international search report
16 JUN 2004

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703)305-3230

Authorized officer
Teresa Strzelecka Janice Ford for
Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/18628

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-30

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-30, drawn to a kit for identifying an origin of a tumor, the kit comprising two probes which can detect expression products of two genes from Table 3 and a method of identifying an origin of a tumor by detecting expression levels of at least two genes identified in Table 3.

Group II, claim(s) 31-41, drawn to a method for identifying an origin of a tumor by providing a predictor set for two or more genes, detecting in a tumor an expression level for a gene diagnostic for a tumor, calculating a vector distance from the expression level obtained from the tumor sample to each of the levels in the predictor set, determining the shortest vector distance.

Group III, claim(s) 42-48, drawn to a method for obtaining a predictor set for classifying a sample into one or more classes, by obtaining a value for one or more features for each of plurality of members of each of the classes, determining a Wilcoxon rank score for each of the features and ranking the features by predictive accuracy.

Group IV, claim(s) 49-56, drawn to a method for screening a subject for prostate cancer by detecting a level of expression of at least one gene in a sample of prostate tissue, comparing the value with a level of expression of the gene in a sample from a disease free subject.

Group V, claim(s) 57-64, drawn to a method for screening a subject for ovarian cancer by detecting a level of expression of at least one gene in a sample of ovarian tissue, comparing the value with a level of expression of the gene in a sample from a disease free subject.

Group VI, claim(s) 65, drawn to a method for monitoring the progression of prostate cancer in a subject by detecting a level of expression of at least one gene in a sample of prostate tissue, wherein an increase in the level of expression of the gene over time is indicative of the progression of the prostate cancer.

Group VII, claim(s) 66, drawn to a method for monitoring the progression of ovarian cancer in a subject by detecting a level of expression of at least one gene in a sample of ovarian tissue, wherein an increase in the level of expression of the gene over time is indicative of the progression of the ovarian cancer.

Group VIII, claim(s) 67, drawn to a method of identifying agents for use in treatment of prostate cancer by contacting a sample of diseased prostate cells with a candidate agent, detecting a level of expression of at least one gene in the sample and comparing the levels of expression of the gene before and after addition of the candidate agent.

Group IX, claim(s) 68, drawn to a method of identifying agents for use in treatment of ovarian cancer by contacting a sample of diseased ovarian cells with a candidate agent, detecting a level of expression of at least one gene in the sample and comparing the levels of expression of the gene before and after addition of the candidate agent.

Group X, claim(s) 69 and 70, drawn to a method of inhibiting undesired proliferation of a prostate cell by administering to the cell an effective amount of an agent that can decrease expression of at least one gene.

Group XI, claim(s) 71 and 72, drawn to a method of inhibiting undesired proliferation of an ovarian cell by administering to the cell an effective amount of an agent that can decrease expression of at least one gene.

Group XII, claim(s) 73, drawn to a method for monitoring the efficacy of a treatment with an agent of a subject having prostate cancer by obtaining pre-and post-administration samples from the subject, obtaining expression levels of at least one gene in the samples and comparing the expression levels in pre- and post-administration samples.

Group XIII, claim(s) 74, drawn to a method for monitoring the efficacy of a treatment with an agent of a subject having ovarian cancer by obtaining pre-and post-administration samples from the subject, obtaining expression levels of at least one gene in the samples and comparing the expression levels in pre- and post-administration samples.

The inventions listed as Groups I-XII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical feature of Group I is detection of the expression of two genes in a sample, the special technical feature of Group II is identification of tumor origin by calculating a vector distance from the predictor set to the expression level from the tumor sample and determining the shortest vector distahce, the special technical feature of Group III is identification of a predictor set, the special technical feature of Groups IV and V is screening for prostate cancer or ovarian cancer, respectively, the special technical feature of Groups VI and VII is monitoring the progression of prostate cancer or ovarian cancer, respectively, the special technical feature of Groups VIII and IX is identifying agents for treatment of prostate cancer or ovarian cancer, respectively, the special technical feature of Groups X and XI is inhibiting proliferation of prostate cancer or ovarian cancer cells, respectively, and special technical feature of Groups XII and XIII is monitoring the efficacy of a treatment of prostate cancer or ovarian cancer subject, respectively.